ARTICLE

Central Place Indexing: Hierarchical Linear Indexing Systems for Mixed-Aperture Hexagonal Discrete Global Grid Systems

Kevin Sahr

Department of Computer Science / Southern Oregon University / Ashland / OR / USA

ABSTRACT

Hexagonal discrete global grid systems (DGGSs) with integer spatial indexes are a promising new approach to designing geospatial data structures and location reference systems. Central place indexing (CPI) is a class of multi-precision hierarchical linear spatial indexing systems for pure and mixed-aperture hexagonal DGGSs. Definitions for CPI systems are given both on the plane and on the polyhedral surfaces of geodesic DGGSs, and examples of real-world DGGSs indexed using CPI are described. The semantic advantages of CPI systems are discussed, including their ability to exactly represent their own geometries.

Keywords: discrete global grids, spatial indexing, geocoding, spatial data structures

RÉSUMÉ

Les systèmes de grilles globales discrètes (DGGS) hexagonales faisant appel à des indices spatiaux entiers offrent une nouvelle méthode prometteuse pour la conception de structures de données géospatiales et de système de géoréférencement. Les systèmes d'indexation de lieux centraux (*central place indexing* — CPI) constituent une catégorie de systèmes d'indexation spatiale linéaire hiérarchique multiprécision pour les DGGS hexagonaux purs et à mailles mixtes. L'auteur définit les systèmes CPI tant sur la surface plane que sur la surface polyédrique des DGGS géodésique, et il propose des exemples concrets de DGGS indexés à l'aide de systèmes CPI. Il analyse les avantages sémantiques des systèmes CPI, y compris leur capacité de représenter avec exactitude leur propre géométrie.

Mots clés : géocodage, grilles globales discrètes, indexation spatiale, structures de données spatiales

Introduction

The impressive achievements of geospatial computing have been enabled by the development of powerful geospatial location representation systems and data structures. But, as useful and convenient as traditional systems have proven, it is questionable whether they can meet the challenges of global geospatial big data and of truly realizing the Digital Earth vision (Goodchild and others 2012). Researchers (e.g., Dutton 1999; Goodchild 2018; Goodchild and others 2012; Sahr 2011) have argued for the development of data representations based on discrete global grid systems (DGGSs; Sahr, White, and Kimerling 2003) – regular multi-resolution partitions of the sphere into geocoded cells. A substantial body of research (see the survey in Sahr 2011) suggests that DGGS based on regular partitions of

spherical Platonic solids into cells that are primarily hexagonal in shape exhibit superior properties for a wide range of use cases, providing superior representational efficiency, semantic fidelity, and algorithmic tractability. Hexagonal DGGS are already being used in a variety of scientific and commercial settings (e.g., Brodsky 2018; Open Geospatial Consortium 2017; Sahr 2018a; Sahr, Dumas, and Chaudhuri 2015).

Hexagonal DGGS consist of multiple resolutions of hexagonal grids, and the geometric relationship between these resolutions is often defined in terms of their aperture (Bell and others 1983), or the ratio of areas between a hexagonal cell in successively coarser grid resolutions. As discussed in the next section, indexing these cells using hierarchical linear indexes has many advantages, and several researchers have developed such indexing systems for hexagonal

DGGS (e.g., Sahr 2008; Tong, Ben, and Wang 2010; Wang and others 2017; White, Kimerling, and Overton 1992). But these indexing systems have largely been restricted to pure sequences of grid resolutions of aperture 3 or 4 – two of the three central place apertures (Christaller 1966). The grid system of White and others (1992) is the only geospatial grid system to index mixed sequences of all three central place apertures – 3, 4, and 7 – but that grid is not extended to the entire sphere.

This article presents central place indexing (CPI): a uniform multi-precision hierarchical linear indexing system for pure and mixed-aperture hexagonal DGGSs. CPI indexing allows the cells in the indexed DGGS to embody the semantics of either raster/gridded or vector/point location representation, or to be used as efficient and algorithmically tractable data buckets for geospatial sharding and databases. First some background on planar hexagonal grids is given, and then CPI systems are defined on the plane before that definition is extended to polyhedral surfaces. Algorithms are described for the geometric generation of cells in CPI systems, and some examples of implemented CPI systems are discussed. Finally, some useful features of CPI systems are described, such as their ability to exactly represent their own cell geometries.

Background

AN EXEMPLAR: THE SQUARE QUADTREE

It is useful to assign to each cell in a hexagonal DGGS a unique structured integer index. On the plane the canonical example of such an indexing is the aperture 4 square quadtree (Gargantini 1982). Quadtree indexes are hierarchical prefix codes; each digit in a cell index corresponds to a cell at some coarser precision, defined relative to their common hierarchical indexing ancestors. Such indexes implicitly encode both precision (indicated by the number of digits in the index), without the need for metadata, and direction (encoded in the well-defined digit arrangement at each level of precision), as well as defining a locality-preserving total ordering of the cells (i.e., a space-filling curve) at each precision (Mark and Goodchild 1986).

Every quadtree cell is the root of an indexing hierarchy. Let the precision q indexing footprint of a precision p root cell, where $q > p$, be the set of precision q cells whose p-digit prefix is the index of the precision p root cell. In a square quadtree, each cell's indexing hierarchy is equivalent to that cell's spatial hierarchy (Kiester and Sahr 2008), consisting of the cells at finer precisions that are contained within, or overlap, the root cell. The cells at each precision in the indexing footprint/spatial hierarchy of a cell form a partition of that cell. Square quadtree indexes naturally provide multi-precision quantification (Sahr 2013); if a point location is quantized at some desired precision p by assigning to it the index of the precision p cell that contains the point, then the precision p index contains, as prefixes, the correct

quantification indexes for all coarser precisions. The index of the precision q quantification of the point, where $q < p$, is simply the first q digits of the precision p index.

The term resolution, rather than precision, is often used in this context. Resolution refers to the minimum difference between two location values that can be distinguished by the system (Csillag 1991), and therefore seems more appropriate for discussing fully specified DGGSs (per Sahr and others 2003), where each cell corresponds to a specific region on the surface of the earth. But when indexing systems are defined independent of particular DGGSs, as is being done here, the most that can be said is that each digit in the hierarchical index supplies a certain number of additional bits that may be used to distinguish cells at that precision. The cells of whatever DGGS will be indexed will correspond to specific geospatial regions, but the resolution of those regions is not determined solely by the precision of the cells.

Quadtree indexes have multiple fully equivalent semantics, which can be freely moved between in conceptualizing quadtree construction and use. For example, indexes for the cells at a given precision can be algorithmically constructed in at least three ways, and the choice made has no bearing on the approach taken to designing algorithms that make use of the resulting indexes. Quadtree indexes can be constructed top-down, by recursive subdivision, beginning with a single large square that covers the region of interest. But they can also be constructed bottom-up, by starting at the desired precision and recursively aggregating the square cells in groups of four to form larger squares, until a square is created that covers the domain. Equivalently, a recursive fractal space-filling curve can be constructed at the desired precision, and transformations that map square cells at a particular precision to and from such curves effectively assign quadtree indexes to those cells, without conceptual recourse to other grid precisions.

INDEXING PLANAR HEXAGON GRIDS

In the case of hexagonal cells, it is impossible to partition a hexagon exactly into smaller hexagons or to aggregate smaller hexagons to form a larger hexagon; hexagonal hierarchies do not have the multiple equivalent semantics of the quadtree. But a multi-precision sequence of aligned hexagon grids can be constructed, where the centre points of hexagons at a given precision are also centre points of hexagons at the next finer precision (and hence at all finer precisions; Bell and others 1983). The apertures that form aligned multi-precision hexagon grids are generated by the formula $i^2 + ij + j^2$, where i and j are arbitrary nonnegative integers (Dacey 1965). The smallest three such apertures are 3, 4, and 7, known as the central place apertures (Figure 1), after Christaller (1966), who argued that ideal human settlement patterns form aperture 3, 4, and 7 hierarchies. The smaller branching factors of these three
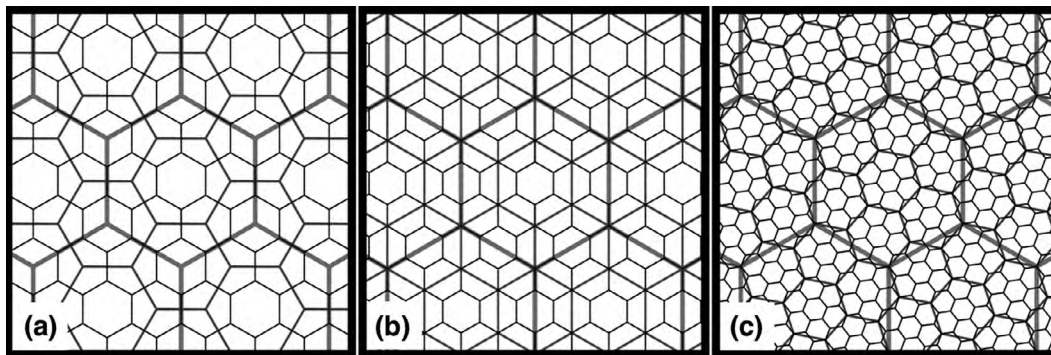
Kevin Sahr

Figure 1. Three precisions of aligned multi-precision hexagonal grids using the three central place apertures: (a) aperture 3, (b) aperture 4, and (c) aperture 7

apertures also make them attractive for creating hierarchical indexing structures.

Hierarchical prefix codes can be created for the cells of a single-precision hexagon grid bottom-up, analogously to a quadtree, by aggregating the cells into compact groups of 3, 4, or 7 hexagons and then recursively aggregating those groups into larger groups, until only a single group remains (see Figure 2). The groups at each level of aggregation form the nodes of a single precision of the corresponding aperture sequence of grids. Indexes are formed by consistently assigning a digit per precision, with the digit base traditionally determined by the aperture (Burt 1980; Gibson and Lucas 1982). Tesseral arithmetics (Bell and Holroyd 1991; Diaz and Bell 1986) can be defined on these indexes, with common vector operations such as addition and scaling using very efficient per-digit integer calculations or table lookups.

Computing with square grids traditionally uses two-dimensional integer coordinate systems with orthogonal axes. Hexagon grids have three natural axes spaced 120° apart, as illustrated in Figure 3. Any two of these axes are sufficient to identify each hexagon uniquely, using a 2-tuple of integers, just as with square grid coordinate systems. Hierarchical space-filling curves defined on two-di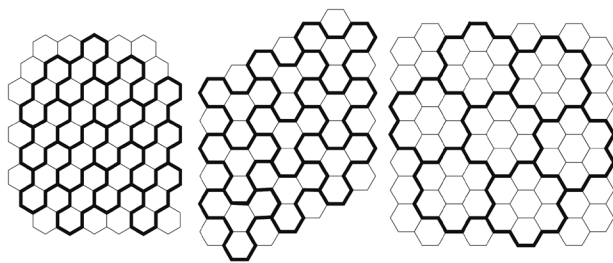mensional square grids can be constructed directly on hexagonal grids, with common curves such as z-order and Hilbert's curve – which generate an aperture 4 spatial and indexing hierarchy on square grids – similarly inducing an aperture 4 hexagonal sequence of precisions (see Figure 4).

As with a square quadtree, these same hexagonal indexing hierarchies can be viewed as top-down constructions. But unlike squares, it is not possible simply to partition hexagonal cells into smaller hexagons. Hierarchical prefix codes can be created by starting with a hexagon at the coarsest precision, and then assigning to it a set of hexagons at the next finer precision that will be its indexing children. While the indexing children of a parent cell in a square quadtree are fully determined by the spatial hierarchy, the assignment of child cells to parents in hexagonal hierarchies is ambiguous; a coarse-precision hexagon wholly contains some hexagons at the next finer precision, but there are other hexagons that it only partially overlaps,
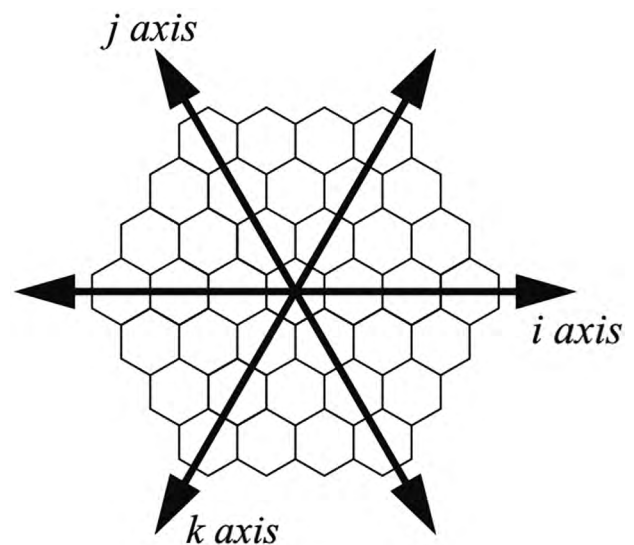

Figure 2. Example recursive planar hexagonal aggregation groupings with apertures 3, 4, and 7, respectively (see also Figures 11 and 12 in the next section)

Figure 3. The three natural axes of a hexagonal grid, spaced 120° apart
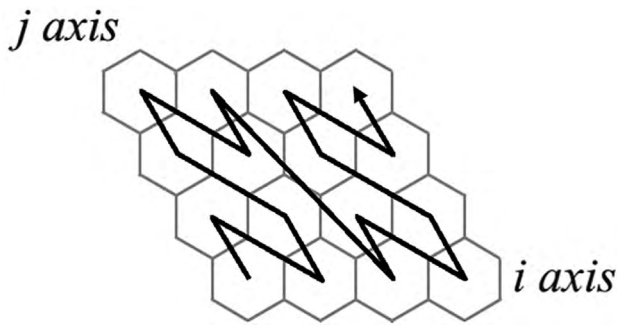
Figure 4. A *z*-order space–filling curve defined on a two-dimensional hexagonal coordinate system

which could potentially be indexed by a cell adjacent to the parent under consideration. One solution is to note, as done above, that a bottom-up approach results in groupings that correspond to the nodes at single precisions of the corresponding aperture sequence of grids, and the hierarchical prefix codes formed bottom-up trace an indexing hierarchy through these grids. These same groupings can be constructed top-down to unambiguously assign child cells to parents. This idea will be developed further in the next section.

Recursive aperture 7 grouping of hexagons best approximates a hexagonal shape across all precisions and has therefore received the most attention on the plane. One efficient digit assignment for each aperture 7 unit is generalized balanced ternary (GBT; Gibson and Lucas 1982), a generalization of one-dimensional balanced ternary addressing (Knuth 1998, 190–92) to the three axes of a hexagon grid, as illustrated in Figure 5.

Aggregation-based hexagonal grid indexing, whether constructed bottom-up or top-down, can be used to index a
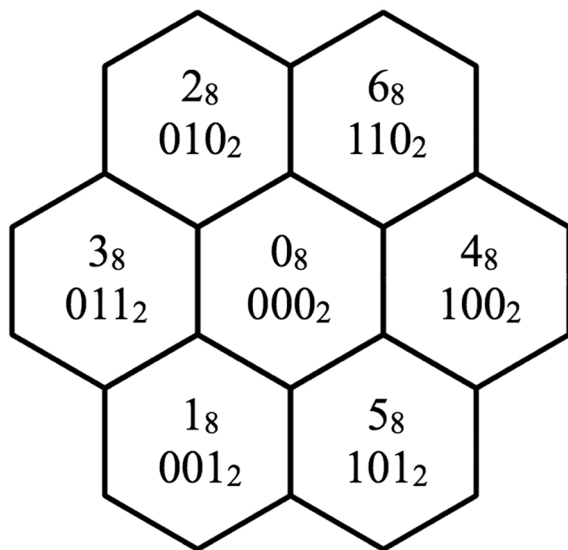
single-precision raster grid hierarchically. White and others (1992) used aggregation-based indexing to create a hexagonal grid system with a user-specified sequence of apertures 3, 4, and 7 on a single hexagonal face of a truncated icosahedron, projected to create cells on the sphere using an equal-area projection. Since the indexing hierarchy is anchored to a specific geospatial region – a polyhedron face – the coarsest grid precision is fixed, making a top-down construction much more straightforward to define and implement.

CPI System Definition

The topology and location coding of CPI systems are defined on *ideal CPI manifolds*. An ideal CPI manifold is a triangulated two-dimensional surface, with equal-length edges connecting adjacent vertices to create planar triangle faces, and where each of the vertices has a valence, or number of triangular faces for which it is a vertex, of at most 6. Interesting ideal CPI manifolds include a regular triangular lattice on the plane and triangle-faced regular polyhedra such as the icosahedron.

A topology and indexing defined on an ideal CPI manifold can be applied to arbitrary triangulated two-dimensional manifolds by defining an appropriate mapping between these manifolds and a topologically equivalent ideal CPI manifold. For example, a CPI hierarchy can be constructed on an arbitrary two-dimensional triangulated manifold with unequal edges (such as a triangulated irregular network or a stellated polyhedron) by defining a mapping (e.g., bilinear interpolation) between the irregular triangles adjacent to each vertex and the corresponding equilateral triangles on a topologically equivalent ideal CPI manifold.

Similarly, CPI systems can be created on triangulations of a curved 2D surface by defining a mapping between the curved triangles and an ideal CPI manifold, as is often done when constructing a DGGS. For example, the ISEA3H and ISEA4H hexagonal DGGSs (Sahr and others 2003) use the icosahedral Snyder equal-area (ISEA) projection (Snyder 1992) to map the planar triangular faces of an icosahedron to and from the spherical triangles of a spherical representation of the earth's surface, while the H3 hexagonal DGGS (Brodsky 2018) uses an icosahedral gnomonic projection (Fisher 1943) for the same purpose.

CPI systems can be defined using a top-down approach. Let **M** be an ideal CPI manifold. A *cell* at precision p of a CPI system defined on **M** consists of the following:

(1) A point on **M** that is either a vertex of **M** (if $p = 0$) or introduced through generation of subsequent precisions of the CPI system (if $p > 0$).

(2) The polygonal Voronoi area on **M** associated with that point, defined relative to all other precision p cell points.

(3) A *generator* that specifies the geometry, generator types, and location coding of all precision p+1 cells

Figure 5. GBT digit assignment

Kevin Sahr

that are children of this cell in the indexing hierarchy defined by the location coding system.

Then a *CPI system specification* consists of the following:

(1) A connected set of precision 0 cells, referred to as the system's *base cells*.

(2) A sequence of apertures 3, 4, and/or 7 that define the topology of each finer precision in the system. As discussed below, in the case of apertures 3 and 7 a direction of rotation must also be specified.

CPI INDEX DEFINITION ON A VALENCE-6 VERTEX

If a single base cell is formed on a vertex of **M** with a valence of 6, the resulting Voronoi cell on an ideal CPI manifold will be a planar hexagon.

The *central place children* of a valence-6 precision b cell c_b consist of an appropriately scaled and rotated precision $b+1$ cell centred on c_b and the six adjacent precision $b+1$ cells. In the case of aperture 3, the central place children will have 1/3 the area of c_b and are rotated 30° counter-clockwise or clockwise relative to it (see Figures 6a and 6b, respectively). In the case of aperture 4, the central place children will have the same orientation as c_b and

1/4 its area (Figure 6c). Finally, in the case of aperture 7, the central place children will have 1/7 the area of c_b and will be rotated by approximately 19.1° counter-clockwise or clockwise relative to it (Figures 6d and 6e, respectively). Note that the aperture 3 clockwise and counter-clockwise cases generate central place children that are geometrically identical, though this is not true of aperture 7. However, in both the aperture 3 and 7 cases each rotation direction will generate different indexes (see below).

This process can be applied recursively at precisions $b+2$, $b+3$, . . . (until a desired precision is achieved) by choosing the aperture for each precision given in the CPI system specification. Let the precision p Christaller set of c_b be the resulting set of cells at precision p (where $p > b$). Figures 7 and 8 illustrate Christaller sets generated by some representative pure and mixed aperture sequences. Note that in the case of aperture sequences involving apertures 3 and/or 4 some cells will be generated multiple times.

A hierarchical integer location code (or codes) can be assigned to a precision p cell c_p in the Christaller set of c_b as follows. The location code of c_p will have as its prefix the integer location code of c_b, with a single digit concatenated to it for each precision from $b+1$ to p, inclusive. Since each cell has seven central place children, the digits 0, 1, 2, . . . , 6 are a convenient choice for these additional digits. The assignment
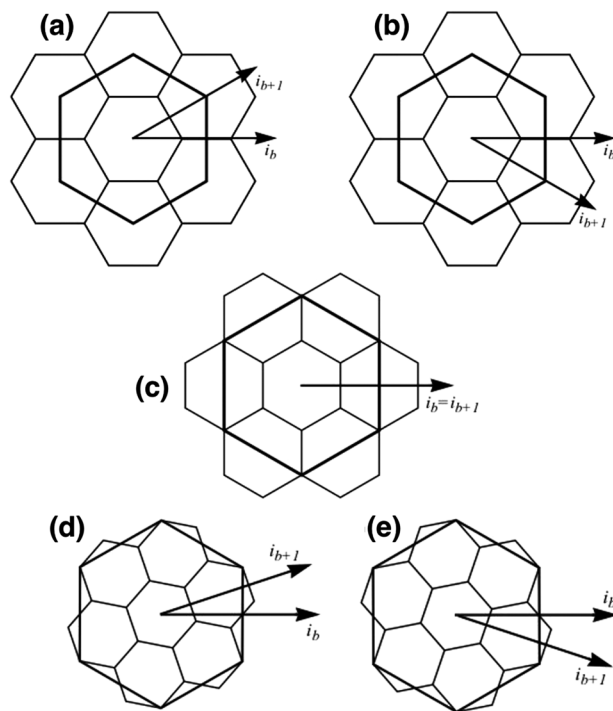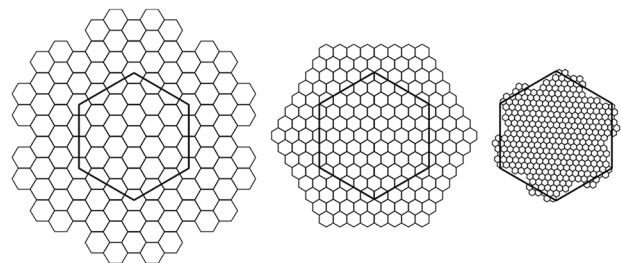

Figure 6. Central place children under all possible central place apertures: (a) aperture 3 with counter-clockwise rotation (3^{ccw}); (b) aperture 3 with clockwise rotation (3^{cw}); (c) aperture 4 (4); (d) aperture 7 with counter-clockwise rotation (7^{ccw}); (e) aperture 7 with clockwise rotation (7^{cw})
Source: Sahr (2013).

Figure 7. Precision $b + 3$ Christaller sets generated by pure central-place aperture sequences: (a) 3^{ccw}, 3^{ccw}, 3^{ccw}, (b) 4, 4, 4, and (c) 7^{ccw}, 7^{ccw}, 7^{ccw}
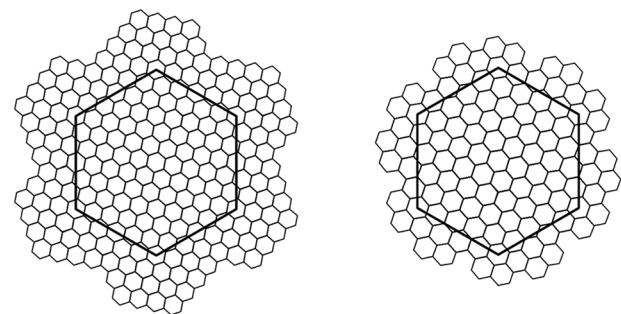

Figure 8. Examples of precision $b + 3$ Christaller sets generated by mixed aperture sequences: (a) 3^{ccw}, 7^{ccw}, 4, (b) 7^{cw}, 3^{ccw}, 3^{cw}

of digits should be geometrically consistent and can be specified by assigning each digit to a child relative to that child's position in the local coordinate system at that child's precision. Let the arrangement given in Figure 9 be the default CPI arrangement. Table 1 lists several useful ways in which these digits might be represented. Note that this arrangement is rotated from the canonical arrangement used in GBT, so that unit vectors in the i, j, and k directions correspond to the binary representation of the octal digits, which facilitates working with the system. A CPI index for c_p is any one of the location codes formed by following a single hierarchical path through the Christaller set of c_b that generates c_p.

A *cell generator* must also be specified for each cell, which indicates which of the cell's seven child cells are actually indexed by that generator cell, along with the cell generator associated with each of the generated children. Thus, generators recursively define subsets of the possible paths through the Christaller sets of a given root cell.

Since each cell in a CPI system has at most seven children, let the generator for a cell c be specified as a string of seven values $g_0 g_1 g_2 g_3 g_4 g_5 g_6$, where each g_i specifies the generator

associated with the i-digit child of c (as defined in the previous paragraph).

Then the $\mathbf{C}^7$ generator is

$$\mathrm{C}^7 = \mathrm{C}^7 \mathrm{C}^7 \mathrm{C}^7 \mathrm{C}^7 \mathrm{C}^7 \mathrm{C}^7 \mathrm{C}^7.$$

A CPI system can now be completely specified on a valence-6 ideal CPI manifold as follows. Let the base cells be centred on the lattice vertexes and assign the $\mathbf{C7}$ generator to each base cell. Then choose a sequence of one or more apertures: 3 (counter-clockwise and/or clockwise), 4, and/or 7 (counter-clockwise and/or clockwise).

GEOMETRIC GENERATION OF CPI CELLS

Each precision p of the Christaller set of a precision-b cell c_b (where $b < p$) is created using a consistent scaling and rotation across all cells at that precision, relative to the previous precision $p-1$. This means that all precision p cells lie on a regular hexagonal grid that is scaled by the product of the precision $b+1$ to p scaling factors and rotated by the sum of the precision $b+1$ to p rotations about the origin (as illustrated in Figure 6).

Let $\mathbf{A}$ be a central place aperture sequence, and let a_A be the number of central place apertures of type a that occur in $\mathbf{A}$. Then scaling factor sp for precision p relative to precision 0 in a CPI system with aperture sequence $\mathbf{A}$ is given by

$$s_p = \frac{1}{\sqrt{3}^{\left(3_A^{\mathrm{ccw}} + 3_A^{\mathrm{cw}}\right)} \times 2^{4_A} \times \sqrt{7}^{\left(7_A^{\mathrm{ccw}} + 7_A^{\mathrm{cw}}\right)}}$$

The rotation angle dp for precision p relative to precision 0 in a CPI system with aperture sequence A is given by

$$d_p = \left((3_A^{\mathrm{ccw}} - 3_A^{\mathrm{cw}}) \times 30° \right) + \left((7_A^{\mathrm{ccw}} - 7_A^{\mathrm{cw}}) \times \arcsin\left(\sqrt{\frac{3}{28}} \right)° \right).$$

NORMALIZED CPI SYSTEMS

Hierarchical generation using $\mathbf{C}^7$ generators under aperture sequences that include apertures 3 and/or 4 will result in the assignment of multiple CPI indexes to some of the cells. If the system is being used to geocode a point location, then a multi-precision quantization (Sahr 2013) can
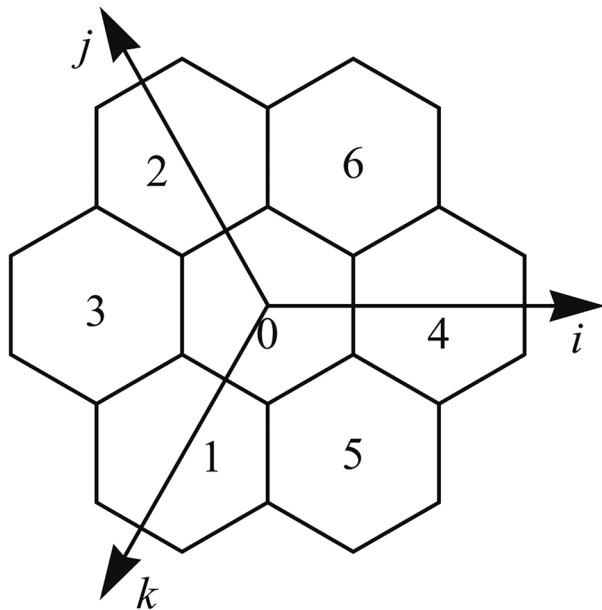

Figure 9. Digit assignment at successive precisions, defined relative to the three-dimensional hexagonal coordinate system at that precision

Table 1. Three representations of direction in CPI systems

Octal	0	1	2	3	4	5	6
(i, j, k)	(0, 0, 0)	(0, 0, 1)	(0, 1, 0)	(0, 1, 1)	(1, 0, 0)	(1, 0, 1)	(1, 1, 0)
Binary	000	001	010	011	100	101	110

Note: Note the equivalence between (i, j, k) coordinate components and digits in the corresponding binary representation.

be created by choosing, for each precision p, a digit that corresponds to the cell that contains the point location at that precision (see example in Figure 10).

It is often useful to specify generators that index each cell uniquely. This includes use cases where the cells represent raster pixels or data structure buckets, as well as cases where generators are used to hierarchically generate cell geometries. Let a *normalized* CPI system be one in which each cell has only one valid CPI index. Note that aperture 7 systems always assign a unique CPI index to each cell, and therefore are always normalized.

Central place generators can be defined for an aperture a of 3 or 4 that generate a corresponding number of children: the central (0-digit) central place child, and $a-1$ of the other six central place children, where each such generator can tile an ideal planar CPI manifold. Figures 11 and 12 illustrate the single generators (unique under reflections and 60° rotations), that tile the plane with consistent orientation, for counter-clockwise aperture 3 and aperture 4 precisions respectively, along with the corresponding generator string representations. A child cell that is not generated is indicated in the string representation by replacing a generator specification with the corresponding digit 0–6. **G** is used to indicate an arbitrary generator type.

The precision p indexing footprint of a cell in an aperture 3 or 4 normalized CPI system is a subset of that cell's precision p Christaller set. Only in the case of aperture 7 is the indexing footprint of a cell always equal to its Christaller set at every precision. Figure 13 illustrates the assignment of a cell index for a particular normalized CPI system.
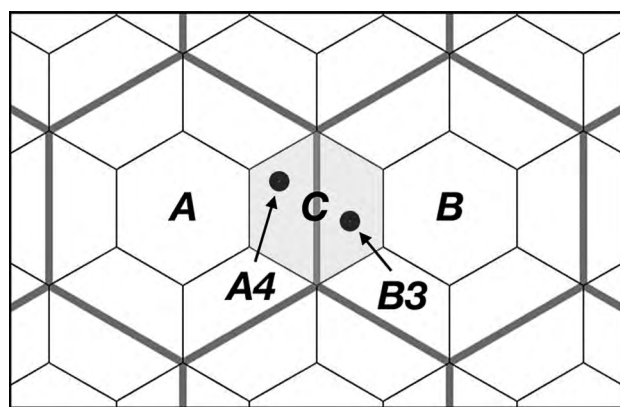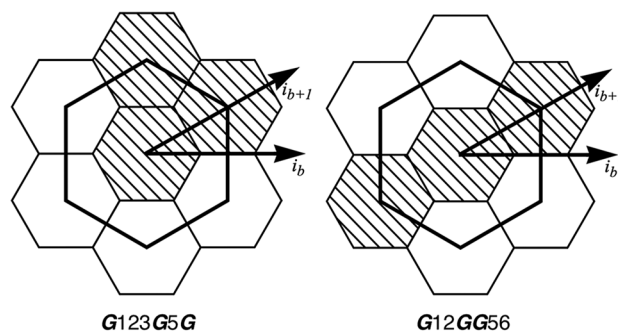

Figure 11. Generator types that tile the plane with consistent orientation for counter-clockwise aperture 3.
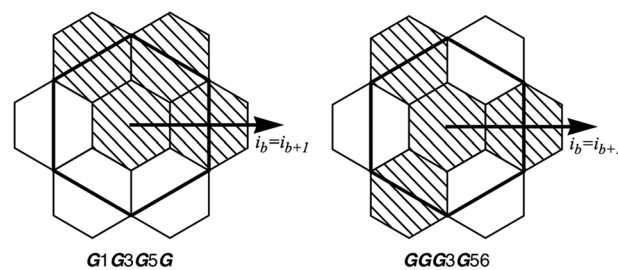
Figure 12. Generator types that tile the plane with consistent orientation for aperture 4.

Figure 10. Two vector locations quantized into precision 2 of an aperture 4 grid system. At precision 2 the points quantize to the same cell **C**. But at precision 1 they quantize to two different cells, with indexes **A** and **B**, respectively. A multi-precision quantization would assign each point an index based on the corresponding precision 1 base cell, yielding indexes **A4** and **B3** respectively. This approach effectively indexes the subregions formed by the intersections of multiple-precision cells.

Figure 13. An example of using a normalized CPI system to assign a unique index to a cell: (a) a cell **C** at precision 3 of a grid system with aperture sequence 3^{ccw}, 4, 7^{ccw}; (b) the precision 0 base cell **A** with generator G123G5G indexing precision 1 child cell **A4**; (c) the precision 2 cell **A44** indexed using generator GGG3G56; (d) the precision 3 cell **C** assigned index **A445** using the generator C^7.

Single generators that tile the plane using multiple orientations can also be used, as well as combinations of different generators that tile the plane. For example, the aperture 3 generators given in Figure 11 produce indexing footprints that do not have hexagonal symmetry. The aperture 3 hexagon tree (A3HT; Sahr 2008) is a normalized aperture 3 CPI system that generates indexing footprints with full hexagonal symmetry. It uses two generator types, the open generator $\mathbf{A} = \mathbf{B}123456$ and the closed generator $\mathbf{B} = \mathbf{B}AAAAAA$ (see Figure 14).

DEFINITION ON VERTICES WITH VALENCES LESS THAN 6

Note that it is impossible to tile the sphere completely with hexagonal cells; each precision of a hexagonal geodesic DGGS must contain a fixed number of non-hexagonal cells. For example, hexagonal DGGS based on the icosahedron usually have, at each precision, 12 pentagonal cells centred on the 12 valence-5 vertices of the icosahedron. CPI hierarchies can be constructed on triangulated manifolds that include vertices with valences less than 6 by embedding the lower-valence vertices onto a connected set of valence-6 vertices, and then assigning to the lower-valence vertices generators that do not generate all of the subsequences of the complete valence-6 Christaller set.

For example, the **C6** generator, which generates the complete Christaller set for a valence-5 vertex, can be designated as
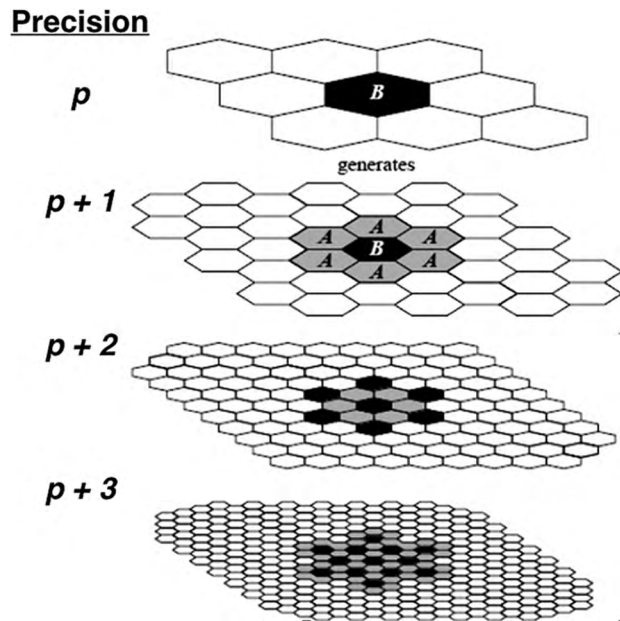
$$C^6 = C^6C^7C^73C^7C^7C^7.$$

Figures 15 and 16 depict the $\mathbf{C^6}$ generator applied to a single base cell with a variety of aperture sequences.

In defining the $\mathbf{C^6}$ generator, any single subsequence can be chosen for non-generation; the choice of subsequence 3 in our definition above is arbitrary. The specific subsequence for deletion can be chosen based on the needs of a particular grid construction, and in any event the same Christaller set cells can always be re-indexed with a different non-generated subsequence by performing a 60° rotation on the subsequences that lie geometrically in between the current and desired orientations. This also enables the construction of multiple valence-6 planar embeddings of portions of a non-valence-6 manifold for the construction of algorithms on the plane.

As with the $\mathbf{C^6}$ generator, the $\mathbf{C^5}$ generator, which generates the complete Christaller set for a valence-4 vertex, can be defined by not generating any two of the subsequences of the full valence-6 Christaller set. Choosing digits 3 and 4 gives the definition

$$C^5 = C^5C^7C^734C^7C^7.$$

A similar approach can be used to create the Christaller sets of vertices with lesser valences.

Just as in the case of valence-6 vertices, note that in aperture sequences that include aperture 3 and/or 4, some of the cells in the Christaller set will be geometrically generated (and indexed) multiple times. Generators that assign
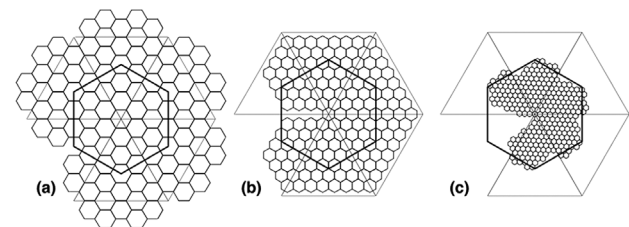
Figure 15. Precision $b + 3$ Christaller sets generated by pure central–place aperture sequences on a valence 5 base tile: (a) 3^{ccw}, 3^{ccw}, 3^{ccw}, (b) 4, 4, 4, and (c) 7^{ccw}, 7^{ccw}, 7^{ccw}.

Figure 14. Four precisions of indexing footprints of a precision p closed (generator type **B**) A3HT base cell. *Source:* Adapted from Sahr (2008) with permission from Elsevier © 2008.
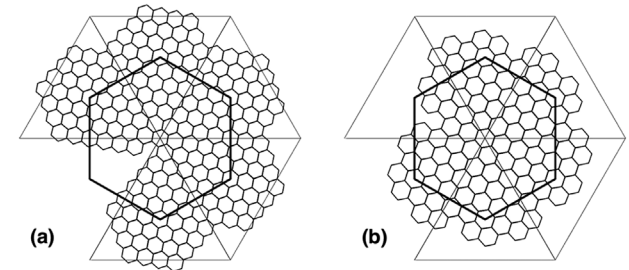

Figure 16. Examples of precision $b + 3$ Christaller sets generated by mixed–aperture sequences on a valence–5 base tile: (a) 3^{ccw}, 7^{ccw}, 4, (b) 7^{cw}, 3^{ccw}, 3^{cw}.

unique cell indexes and that maintain at least the rotational hexagonal symmetries of the base cells (e.g., C^7 for aperture 7, and the A3HT for aperture 3) can be used directly with vertices of any valence, simply by not generating the appropriate number of subsequences under these generators, as described above. Generators that lack hexagonal symmetry must be adapted to the particular symmetries associated with the underlying manifold.

PLANAR CPI ALGORITHMS

A number of planar algorithms for CPI systems are given in Sahr (2010), including forward and inverse quantization, 60° rotation, addition, subtraction, neighbour finding, and metric distance. As with all tesseral arithmetics (Diaz and Bell 1986), discrete integer operations on CPI indexes can be defined efficiently as operations on individual digits, where the per-digit operation is defined using either of two basic approaches: as rule-based operations on groups, or as table lookups. Both approaches involve a small number of equivalently primitive computer operations: integer arithmetic in the case of the group-theoretic rule-based approach and memory accesses in the case of a table-lookup approach. The per-digit operations for basic algorithms are constant-time, and therefore the complexity of most discrete CPI algorithms is $O(p)$, where p is the precision of the CPI operands.

Designing CPI–Indexed DGGSs

CPI defines a class of indexing systems independent of particular DGGSs; the same CPI system could be used to index DGGSs designed for very different use cases. The design of a DGGS often occurs in parallel with the design of the CPI system, and both are dictated by the requirements of the use case. Much research is still required to better understand the optimal hexagonal DGGS and CPI design choices for various use cases. In particular, a full accounting is still needed of the design implications of the semantics of different CPI/DGGS apertures and aperture sequences. The author has participated in the design and implementation of several real-world CPI-indexed hexagonal DGGSs, and the evolution of hexagonal DGGS represented by four examples can help illustrate some of the issues involved. Table 2 summarizes the CPI indexing design choices made for these systems, all of which are defined on an icosahedron.

The icosahedral Snyder aperture 3 hexagon (ISEA3H) DGGS (Sahr, White, and Kimerling 2003) was proposed as a straw man "attempt to construct a good general-purpose Geodesic DGGS," though that analysis implicitly focused on the requirements of the primary use case of that time: representing gridded single-resolution data sets for scientific analysis. Aperture 3 was chosen primarily because, amongst the three central place apertures, it provides the most gradual change in cell area between precisions, giving end users a greater number of grid resolutions from which to choose for gridding their data. But mixed-aperture CPI systems provide many more grid resolution options, making them superior to pure aperture 3 grids for these use cases.

In 2003 the first implementation of the ISEA3H was released as part of the open source program **DGGRID** (Sahr 2018b), developed for the US Environmental Protection Agency to generate DGGSs for survey sampling and gridded data analysis. **DGGRID** generated pure aperture 3 and 4 grids, internally indexed using a pyramid addressing approach (Sahr 2008), with each precision indexed using a set of two-dimensional integer coordinate systems defined on quadrilaterals formed by adjacent pairs of icosahedral faces.

The icosahedral aperture 3 hexagon tree (iA3HT; Sahr 2008) is a CPI system for the ISEA3H and other pure aperture 3 DGGSs. An iA3HT/ISEA3H prototype was successfully implemented for the Canadian Space Agency as a thin layer atop **DGGRID**. The iA3HT base cells are formed by a single unindexed aperture 3 precision (geometrically a truncated icosahedron). A3HT closed (see Figure 14) and open generators are used on pentagonal and hexagonal base cells, respectively. Figure 17 illustrates the indexing

Table 2. Example CPI-indexed DGGSs

Grid	Base cells	Generating method	Aperture sequence
DGGRID aperture 4 INTERLEAVE	precision 0 (pentagons)	*Z*-order space-filling curves	444...
iA3HT	precision 1	open (hexagon) and closed (pentagon) A3HT generators	$3^{ccw}3^{cw}3^{ccw}3^{cw}$...
Superfund_500m	precision 1	specialized generators (see Sahr and White 2010)	$443^{ccw}3^{cw}3^{ccw}3^{cw}$... (15 aperture 3 precisions in total)
H3	precision 2 (CPI43 on Dymaxion icosahedral orientation; Fuller 1975)	generators C^7 (hexagon) and C^6 (pentagon) with 1-digit subsequence removed	$43^{ccw}7^{ccw}7^{cw}7^{ccw}7^{cw}$... (15 aperture 7 precisions in total)

footprints of the base cells for the first six precisions of the iA3HT.

DGGRID includes an aperture 4 CPI index type named INTERLEAVE, which uses a *Z*-order space-filling curve to form the CPI cell indexes. The base cells are the 12 pentagonal vertex cells of the icosahedron, which are the origins of the 2D hexagonal coordinate systems used by **DGGRID**. Given a cell's 2D coordinates, the cell's index is created by interleaving the bits of the *i* and *j* coordinates (see Figures 4 and 18).

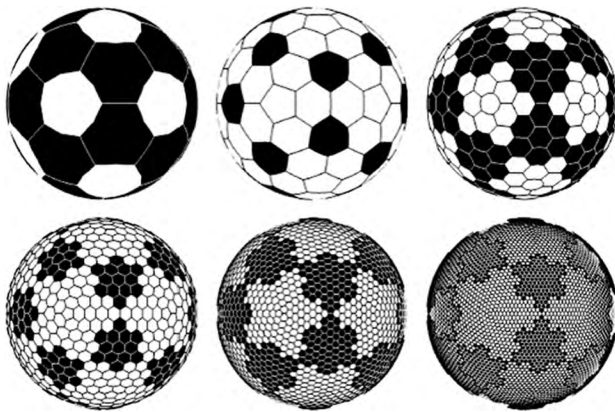The superfund_500m DGGS (Sahr and White 2010) was developed for the US Environmental Protection Agency

Figure 17. The first six precisions of indexing footprints for iA3HT base cells on an ISEA3H DGGS.
Source: Reprinted from Sahr (2008) with permission from Elsevier © 2008.

Superfund Atlas program. An intercell spacing of 500 m was required to balance the requirements of different stakeholders; this could not be achieved with sufficient accuracy using a pure–aperture grid system. A mixed aperture 4 and 3 grid sequence (see Table 2) was used to achieve the required spacing. The base cells are formed by a single un-indexed aperture 4 precision. In order to leverage the existing **DGGRID** software base, a set of 10 CPI generators was used to generate normalized indexing footprints that match the set of quadrilateral 2D coordinate systems used internally by **DGGRID**.

The open source H3 indexing system (Brodsky 2018) was designed for Uber Technologies' use in marketplace analysis and optimization. It is a CPI43 system (Sahr 2013); it uses initial apertures 4 and 3 to encode the icosahedral face center points and algorithmically useful symmetries of the icosahedron. The CPI43 cells are the base cells for a sequence of aperture 7 precisions; these are indexed with C^7 and C^6 generators, applied to hexagonal and pentagonal base cells respectively (see Figure 19). Among the three central place apertures, aperture 7 generates the most "hexlike" of indexing footprints, and these footprints most closely approximate the cells in the corresponding spatial hierarchy. As previously discussed, aperture 7 has unambiguous cell indexing (the indexing footprints and Christaller sets of any cell are equal at all precisions), so it is also the most efficient at utilizing the default CPI octal encoding, with three bits per digit. This makes it a good choice for the efficient and scalable processing of gridded big geospatial data.

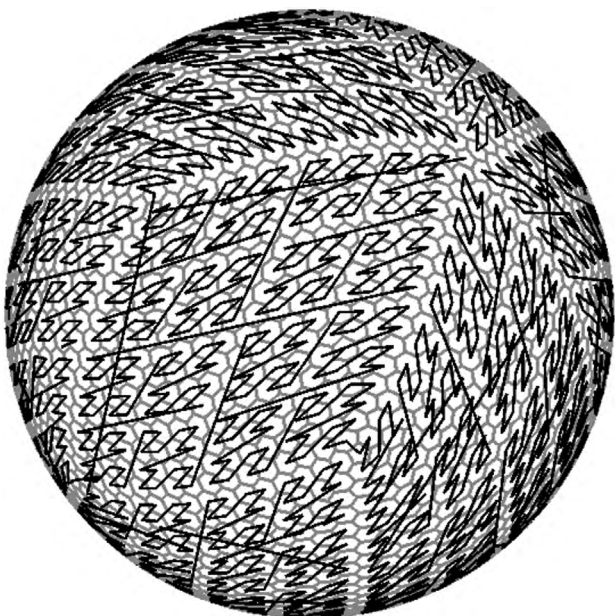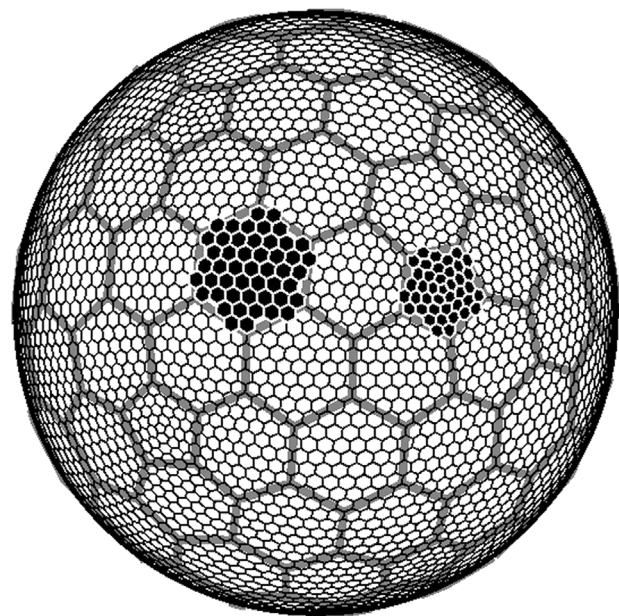**Figure 18.** Precision 4 of a **DGGRID** aperture 4 INTERLEAVE system with *z*-order CPI indexing.

Figure 19. H3 base and precision 2 cells, with the precision 2 indexing footprints of a single pentagonal and hexagonal base cell indicated.

Kevin Sahr

Some Useful CPI Constructions

EXACT REPRESENTATION AND SUBSTRATE GRIDS

In aligned hexagonal grid systems, the centre point of each cell is also a cell centre point at all finer precisions, and CPI systems assign to that centre point the address of the root cell followed by an infinite number of zeros. The location and index of a cell centre point are exactly specified for all finer precisions as soon as it is introduced. Rather than a potentially infinite string of 0s, an octal 7 digit can be used to indicate that all finer precision digits of an index will be 0, allowing us to represent many points with infinite precision, using only a finite number of digits. For example, the centre point of a precision p cell with CPI index P is exactly represented by the index $P7$.

The exact representation of geometric information in a tractable form, such as cell centre points represented as CPI indexes, potentially enables us to store and manipulate those geometric data more efficiently. Algorithms using exact integer location representations can ameliorate the often problematic divergence of numeric computations from actual geometric semantics. Exact representation is not possible, without metadata, using either a floating-point location representation, which is implicitly approximate, or a square quadtree, the cells of which lack a central cell at all precisions.

As discussed in Sahr 2013, the vertices of each cell in a precision p grid are the centre points of cells at an aperture 3 precision $p+1$ grid, as illustrated in Figure 20a. If the aperture sequence of the grid system includes an aperture 3 precision q, where $q > p$, then the vertices of all cells at precision p have an exact representation at precision q. Alternately, an aperture 3 substrate grid – a potentially temporary grid precision lying outside the regular aperture sequence – can be introduced to represent and manipulate the vertices on an as-needed basis. For example, the algorithm used to generate cell boundaries in H3 (Uber Technologies Inc. 2018) uses an aperture 3 substrate grid approach. Similarly, an aperture 4 substrate grid (Figure 20b) can exactly represent the midpoints of each cell edge. And an aperture 7 substrate grid (Figure 20c) introduces exactly represented points in the interior of each cell.

Let the precision-p geometry set be the set of all centre points, vertices, and edge midpoints of cells at precision p. Adding both aperture 3 and 4 substrate grids (in either order) creates a substrate grid that exactly represents all of the points exactly represented by either substrate grid aperture. Given a precision p grid, an aperture sequence 34 substrate grid is a precision $p+2$ grid that exactly represents all points in the precision p geometry set (see Figure 21).

Recall that any CPI grid precision p, where $p > 0$, will be generated from the next coarser precision $p - 1$ grid using either aperture 3, 4, or 7. As illustrated in Figure 22, the points in the coarser precision $p - 1$ geometry set all
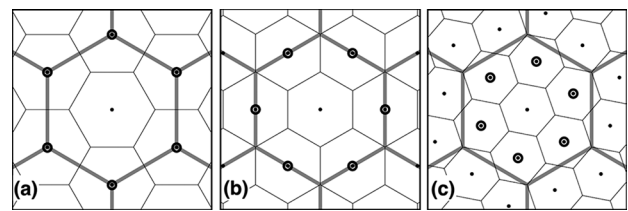

Figure 20. Exact location representations introduced by single precision substrate grids: (a) aperture 3 represents cell vertices, (b) aperture 4 represents the midpoints of cell edges, and (c) aperture 7 represents points in the cell interior.
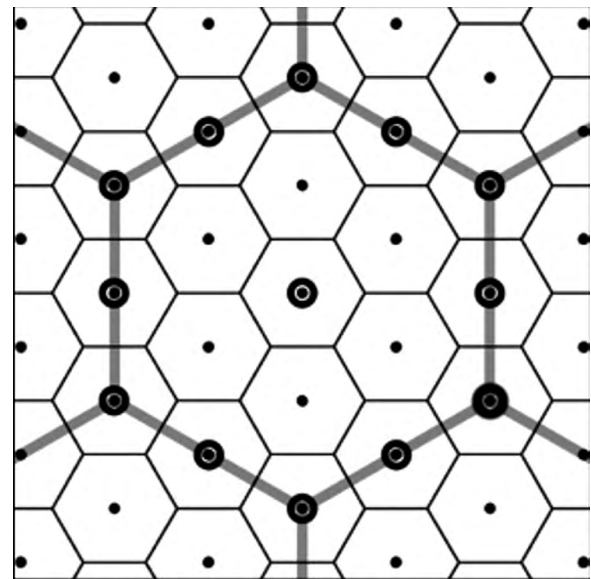
Figure 21. A substrate grid created by adding aperture 3 and aperture 4 to a coarse-precision grid exactly represents the centre points, vertices, and edge midpoints of all of the coarse-grid-precision cells.
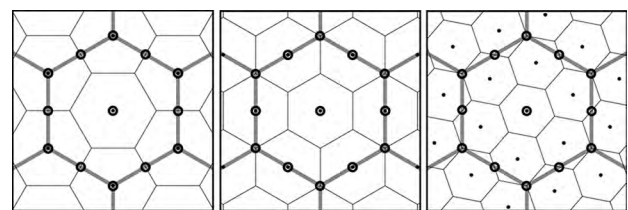
Figure 22. The cell centre points, vertices, and edge midpoints of a coarse-precision grid all coincide with cell centre points, vertices, or edge midpoints in the next-finer-precision grid using aperture 3, 4, or 7, respectively.

coincide with points in the geometry set of the next finer precision p, regardless of the aperture; the precision $p - 1$ geometry set is always a subset of the precision p geometry set. Let **A** be an arbitrary aperture sequence, and let p_{MAX}

be the finest precision in **A**. As discussed above, adding an aperture sequence 34 substrate grid to p_{MAX} will introduce exact representations of the precision p_{MAX} geometry set. Since the precision $p_{MAX} - 1$ geometry set is a subset of the precision p_{MAX} geometry set, the substrate grid also exactly represents the precision $p_{MAX} - 1$ geometry set, and this is true, by induction, for all grid precisions. Thus an aperture sequence 34 substrate grid added to the maximum precision of a CPI system will exactly represent the cell centre points, vertices, and edge midpoints of every grid precision in that system, allowing users to represent and manipulate the geometry of all of these grids using exact integer CPI indexes.

LEAST COMMON SUBSTRATE GRIDS

Let A_1 and A_2 be two CPI aperture sequences defined on the same CPI manifold. A higher-precision common substrate grid can be constructed using an aperture sequence $D = A_1A_2$; the centre points of all cells at all precisions of the A_1 and A_2 grid systems are exactly represented in the precision **D** grid, and each cell in the precision **D** grid can be assigned cell indexes with a prefix in either A_1 or A_2.

Recall that the grid produced by an aperture sequence **A** is completely specified geometrically by the number of each type of aperture that occurs within it, where a_A is the

number of central place apertures of type a (including rotation, as applicable) that occur in **A**. Thus, in constructing a common substrate grid, precisions with apertures that are redundant for the two aperture sequences A_1 and A_2 can be removed from **D**. Let **D'** be an aperture sequence such that $a_{D'} = \max(a_{A1}, a_{A2})$ for each aperture type a. Let the grid produced by **D'** be designated the least common substrate grid, because it is the coarsest precision substrate grid common to both A_1 and A_2. An example of this construction on the sphere is given in Figure 23. Adding an aperture sequence 34 substrate grid to **D'** results in a least common substrate grid that exactly represents the complete geometry set of all cells at all apertures of both A_1 and A_2. This allows us to construct, for any two (or more) disparate CPI systems, a common spatial substrate where the cell geometries of both systems can be efficiently represented and manipulated as integer CPI indexes.

Conclusions

As discussed above, hexagonal DGGS with hierarchical linear indexes have significant advantages for a wide variety of use cases, and mixed-aperture grids provide the greatest known representational and semantic flexibility. CPI is the first systematic approach to the indexing of hierarchical hexagonal DGGS constructed using mixed-aperture central place hierarchies.

Hexagonal DGGS already provide concrete solutions to challenging real-world problems in geospatial computing. This usefulness has led to their implementation and ongoing use in a number of scientific and industrial settings, including the examples of CPI systems described here. A subset of mature DGGS use cases has even begun the process of standardization (Open Geospatial Consortium 2017). It could be argued that the adoption of hexagonal DGGS by end users seems to have outpaced the depth of our understanding of them; much more research is required to better inform effective DGGS design choices and implementations. The author concurs with Ben and others (2018) that much of existing research into hexagonal DGGS "rel[ies] mainly on a research style of induction, guesswork, and experimentation and lack[s] theoretical derivation or proof." And hexagonal DGGS have not been subjected to the kind of rigorous performance comparison testing usually required of geospatial data structures.

But independent of such quantitative analysis, hierarchically indexed hexagonal DGGS provide qualitatively superior solutions for a growing number of geospatial end users. CPI gives these users a tractable approach to efficiently exploiting the full expressiveness of hexagonal DGGS, including their ability to exactly and efficiently represent their own geometries. Further research – including the definition and implementation of more algorithms – will be needed to effectively evaluate the impact of these semantic advantages.

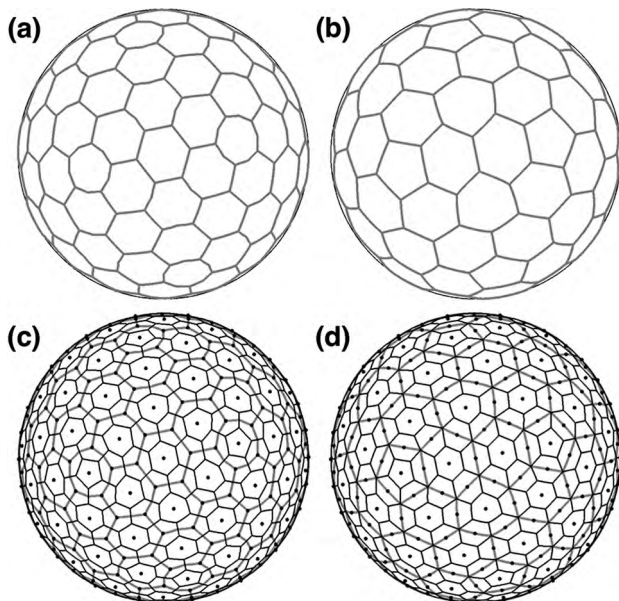**Figure 23.** An example of constructing a least-common-substrate grid on ISEA DGGSs. (a) and (b) are grids generated from the same base cells but with different aperture sequences: $A = 4, 3^{ccw}$ and $B = 3^{ccw}, 3^{ccw}$. (c) and (d) show **A** and **B**, respectively, with a substrate grid generated by the least-common-aperture sequence **D'**, with non-zero aperture components $3^{ccw}_D = 2$ and $4_D = 1$.

Kevin Sahr

Acknowledgements

Author Information

Kevin Sahr is a professor of computer science at Southern Oregon University and director of the Southern Terra Cognita Laboratory (www.discreteglobalgrids.org), which is dedicated to researching and developing discrete global grid systems. E-mail: SahrK@sou.edu.

References

Bell, S.B.M., B.M. Diaz, F. Holroyd, and M.J. Jackson. 1983. "Spatially Referenced Methods of Processing Raster and Vector Data." *Image and Vision Computing* 1(4): 211–20. https://doi.org/10.1016/0262-8856(83)90020-3.

Bell, S.B.M., and F.C. Holroyd 1991. "Tesseral Amalgamators and Hierarchical Tessellations." *Image and Vision Computing* 9(5): 313–28. https://doi.org/10.1016/0262-8856(91)90036-o.

Ben, J., Y. Li, C. Zhou, R. Wang, and L. Du. 2018. "Algebraic Encoding Scheme for Aperture 3 Hexagonal Discrete Global Grid System." *Science China: Earth Sciences* 61(2): 215–27. https://doi.org/10.1007/s11430-017-9111-y.

Brodsky, I. 2018. "H3: Uber's Hexagonal Hierarchical Spatial Index." Available at https://eng.uber.com/h3/.

Burt, P.J. 1980. "Tree and Pyramid Structures for Coding Hexagonally Sampled Binary Images." *Computer Graphics and Image Processing* 14(3): 271–80. https://doi.org/10.1016/0146-664x(80)90056-8.

Christaller, W. 1966. *Central Places in Southern Germany*. Englewood Cliffs, NJ: Prentice Hall.

Csillag, F. 1991. "Resolution Revisited." In *Proceedings of the Tenth International Symposium on Computer-Assisted Cartography: Auto-Carto 10*, ed. American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, 15–29. Bethesda, MD: American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping.

Dacey, M.F. 1965. "The Geometry of Central Place Theory." *Geografiska Annaler: Series B, Human Geography* 47(2): 111–24. https://doi.org/10.2307/490609.

Diaz, B., and S. Bell (eds.). 1986. *Spatial Data Processing Using Tesseral Methods: Collected Papers from Tesseral Workshops 1 and 2*. Reading, UK: Natural Environment Research Council.

Dutton, G. 1999. *A Hierarchical Coordinate System for Geoprocessing and Cartography*. New York: Springer Verlag.

Fekete, G., and L. Treinish. 1990. "Sphere Quadtrees: A New Data Structure to Support Farrell, SPIE, and SPSE, 242–50. Bellingham, WA: SPIE.

Fisher, I. 1943. "A World Map on a Regular Icosahedron by Gnomonic Projection." *Geographical Review* 33(4): 605–19. https://doi.org/10.2307/209914.

Fuller, R.B. 1975. *Synergetics*. New York: MacMillan.

Gargantini, I. 1982. "An Effective Way to Represent Quadtrees." *Communications of the ACM* 25(12): 905–10. https://doi.org/10.1145/358728.358741.

Gibson, L., and D. Lucas. 1982. "Spatial Data Processing Using Generalized Balanced Ternary." In *Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing*, 566–71. Piscataway, NJ: IEEE.

Goodchild, M.F. 2018. "Reimagining the History of GIS." *Annals of GIS* 24(1): 1–8. https://doi.org/10.1080/19475683.2018.1424737.

Goodchild, M.F., H. Guo, A. Annoni, L. Bian, K. de Bie, F. Campbell, M. Craglia, M. Ehlers, J. van Genderen, D. Jackson, A.J. Lewis, M. Pesaresi, G. Remetey-Fülöpp, R. Simpson, A. Skidmore, C. Wang, and P. Woodgate. 2012. "Next-Generation Digital Earth." In *Proceedings of the National Academy of Sciences of the United States of America* 109(28): 11088–94. https://doi.org/10.1073/pnas.1202383109.

Kiester, A.R., and K. Sahr. 2008. "Planar and Spherical Hierarchical, Multi-resolution Cellular Automata." *Computers, Environment and Urban Systems* 32(3): 204–13. https://doi.org/10.1016/j.compenvurbsys.2008.03.001.

Knuth, D. 1998. *The Art of Computer Programming*. Vol. 2. *Seminumerical Algorithms*. Menlo Park, CA: Addison–Wesley.

Mark, D., and M.F. Goodchild. 1986. "On the Ordering of Two-Dimensional Space: Introduction and Relation to Tesseral Principles." In *Spatial Data Processing Using Tesseral Methods: Collected Papers from Tesseral Workshops 1 and 2*, ed. B. Diaz and S. Bell, 179–92. Reading, UK: Natural Environment Research Council.

Open Geospatial Consortium. 2017. "Topic 21: Discrete Global Grid Systems Abstract Specification." Available at http://docs.opengeospatial.org/as/15-104r5/15-104r5.html.

Sahr, K. 2008. "Location Coding on Icosahedral Aperture 3 Hexagon Discrete Global Grids." *Computers, Environment and Urban Systems* 32(3): 174–87. https://doi.org/10.1016/j.compenvurbsys.2007.11.005.

Sahr, K. 2010. "Central Place Indexing Systems." U.S. Patent 9,311,350, filed October 28, 2010, and issued April 12, 2016.

Sahr, K. 2011. "Hexagonal Discrete Global Grid Systems for Geospatial Computing." *Archives of Photogrammetry, Cartography and Remote Sensing* 22: 363–76.

Sahr, K. 2013. "On the Optimal Representation of Vector Location Using Fixed-Width Multi-precision Quantizers." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-4(W2): 1–8. https://doi.org/10.5194/isprsarchives-xl-4-w2-1-2013.

Sahr, K. 2018a. "DGGRID Users." Available at http://www.discreteglobalgrids.org/dggrid-users/.

Sahr, K. 2018b. "DGGRID version 6.4: User Documentation for Discrete Global Grid Generation Software." Available at http://webpages.sou.edu/~sahrk/docs/dggridManualV64.pdf.

Sahr, K., M. Dumas, and N. Chaudhuri. 2015. "The PlanetRisk Discrete Global Grid System." Available at https://webpages.sou.edu/~sahrk/docs/PlanetRiskDGGS.pdf.

Sahr, K., and D. White. 2010. "The Superfund_500m Discrete Global Grid." Available at https://webpages.sou.edu/~sahrk/docs/superfund_500m_020719.pdf.

Sahr, K., D. White, and A.J. Kimerling. 2003. "Geodesic Discrete Global Grid Systems." *Cartography and Geographic Information Science* 30(2): 121–34. https://doi.org/10.1559/152304003100011090.

Snyder, J. P. 1992. "An Equal-Area Map Projection for Polyhedral Globes." *Cartographica* 29(1): 10–21. https://doi.org/10.3138/277-8k88-4882-1752.

Tong, X., J. Ben, and Y. Wang. 2010. "A New Effective Hexagonal Discrete Global Grid System: Hexagonal Quad Balanced Structure."

In *Proceedings of the 18th International Conference on Geoinformatics*, 1–6. Piscataway, NJ : IEEE.

Uber Technologies Inc. 2018. "Walkthrough of h3ToGeoBoundary." Available at https://uber.github.io/h3/#/documentation/core-library/walkthrough-of-h3togeoboundary.

Wang, R., J. Ben, Y. Li, and L. Du. 2017. "The Transform between Geographic Coordinates and Location Codes of Aperture 4 Hexagonal Grid System." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* XLII-2(W7): 161–67. https://doi.org/10.5194/isprs-archives-xlii-2-w7-161-2017.

White, D., A.J. Kimerling, and W.S. Overton. 1992. "Cartographic and Geometric Components of a Global Sampling Design for Environmental Monitoring." *Cartography and Geographic Information Systems* 19(1): 5–22. https://doi.org/10.1559/152304092783786636.