Geometrically Uniform Codes

G. David Forney, Jr., *Fellow, IEEE*

Abstract—**A signal space code $\mathbb{C}$ is defined as geometrically uniform if, for any two code sequences in $\mathbb{C}$, there exists an isometry that maps one sequence into the other while leaving the code $\mathbb{C}$ invariant (i.e., the symmetry group of $\mathbb{C}$ acts transistively). Geometrical uniformity is a strong kind of symmetry that implies such properties as a) the distance profiles from code sequences in $\mathbb{C}$ to all other code sequences are all the same, and b) all Voronoi regions of code sequences in $\mathbb{C}$ have the same shape. It is stronger than Ungerboeck–Zehavi–Wolf symmetry or Calderbank–Sloane regularity. Nonetheless, most known good classes of signal space codes are shown to be generalized coset codes, and therefore geometrically uniform, including a) lattice-type trellis codes based on lattice partitions Λ/Λ' such that $Z^N/\Lambda/\Lambda'/4Z^N$ is a lattice partition chain, and b) PSK-type trellis codes based on up to four-way partitions of a 2^n-PSK signal set.**

Index Terms—**Group codes, trellis codes, geometric codes, Euclidean-space codes.**

I. Introduction

The study of "group codes for the Gaussian channel" was initiated by Slepian [1]. For an excellent recent review of this field, see Ingemarsson [2].

Slepian-type group codes may be characterized as the set of points in real Euclidean N-space $\boldsymbol{R}^N$ that are generated by a group G of orthogonal matrices R, acting on an initial point x_0:

$$\mathbb{C} = \{Rx_0,\ R \in G\}.$$

The size of a Slepian-type group code is finite, and all of its points lie on the surface of a sphere. Slepian-type group codes include M-ary phase-shift-keyed (MPSK) constellations, polyphase codes [3], [4], permutation codes [5], and all linear binary codes with the usual signal space mapping $\{0,1\} \to \{\pm 1\}$ [1], [6].

The group property ensures that the code "looks the same" from any code point. To quote Slepian [1]:

> "Roughly speaking, all words in a group code are on an equal footing: Each has the same error probability [on a Gaussian channel] and the same disposition of neighbors... The set of distances from [any code point] to all other points of the code is the same [for all points]... The maximum likelihood [decision] regions are all congruent."

Manuscript received September 6, 1990. This work was supported in part by the Stanford Center for Telecommunications, Stanford, CA. This work was presented in part at the IEEE International Symposium on Information Theory, San Diego, CA, January 14–19, 1990; and in part at the IEEE Symposium on Information Theory and its Applications, Honolulu, HI, November 1990.

The author is with Motorola Codex, 20 Cabot Boulevard, Mansfield, MA 02048.

IEEE Log Number 9101602.

Signal sets with congruent decision regions are called "completely symmetric" in [7].

Lattice codes have also attracted interest for high-rate applications. A lattice code is a finite subset of points from a lattice Λ, or from a translate $\Lambda + \tau$. A lattice Λ is an infinite discrete subset of $\boldsymbol{R}^N$ that forms an additive group under ordinary vector addition. For an encyclopedic survey of lattices, their properties as packings or coverings, and their group properties, see Conway and Sloane [8].

A lattice translate $\Lambda + \tau$ may be characterized as the set of points in $\boldsymbol{R}^N$ that are generated by the group of translations by elements of Λ,

$$T(\Lambda) = \{t_\lambda \colon x \to x + \lambda,\ \lambda \in \Lambda\},$$

acting on the initial point $x_0 = \tau$:

$$\Lambda + x_0 = \{t_\lambda x_0,\ t_\lambda \in T(\Lambda)\}.$$

A lattice translate therefore has a similar group property which ensures that all of its points "look the same," so that it has similar symmetry properties, such as congruence of decision regions. (Lattice translates are usually not characterized in this manner, because there is an obvious isomorphism between $T(\Lambda)$ and Λ that makes the introduction of $T(\Lambda)$ superfluous.)

For data transmission or quantization, a finite constellation (codebook) is obtained by taking a finite subset of $\Lambda + \tau$. For example, the familiar M-point pulse-amplitude-modulation (PAM) constellation (or the codebook of a uniform scalar quantizer) may be regarded as a finite set of M points taken from a scaled and translated version of the integer lattice $\boldsymbol{Z}$. The properties of such a finite subset are locally the same as those of $\Lambda + \tau$, at least in the interior of the constellation; for example, all interior decision regions are congruent.

This paper studies a more general class of group codes that have similar symmetry properties, called *geometrically uniform codes*. These codes are more general than Slepian-type group codes or lattice codes in two respects.

First, the elements of the generating group G are allowed to be arbitrary isometries, rather than only orthogonal transformations or only translations. When G includes translations, then $\mathbb{C}$ is necessarily an infinite periodic array, called a *regular array*, rather than a finite spherical code. A lattice translate $\Lambda + \tau$ is an example of a regular array. But regular arrays need not be translates of lattices; more generally, regular arrays may be nonlattice crystal structures in N-space, such as are studied in mathematical crystallography [9], [10].

The second respect in which geometrically uniform codes may be more general, which is the principal motivation of this paper, is that $\mathbb{C}$ may be defined as a set of sequences in a possibly infinite-dimensional sequence space. For example, $\mathbb{C}$ may be a trellis code. Trellis codes have become the most popular codes for Gaussian channels, and a large variety of such codes are known (see, e.g., [11]). Our principal result is that practically all known good trellis codes can be characterized as group codes, or *generalized coset codes*, and are therefore geometrically uniform.

If the elements of sequences in a trellis code $\mathbb{C}$ are drawn from an MPSK constellation, then $\mathbb{C}$ is a PSK-type trellis code, which may be regarded as a generalized Slepian-type group code. If elements are drawn from a finite subset of a translate of a lattice, then $\mathbb{C}$ is a lattice-type trellis code, which may be regarded as generalized regular array, or as a crystal in sequence space.

In Section II, we give a brief introduction to isometries and isometry groups, and define geometrical uniformity. In Section III, we note that subgroups of isometry groups induce *geometrically uniform partitions* of geometrically uniform signal sets into geometrically uniform subsets. Geometrically uniform partitions include the kinds of partitions that were used by Ungerboeck [12] in his original paper on trellis codes, as well as the partitions of lattice-type signal sets into subsets corresponding to cosets of a sublattice that were introduced into the trellis coding literature by Calderbank and Sloane [13] (see also [11]).

In Section IV, we show that when a geometrically uniform partition is labeled by an appropriate label group A in an *isometric labeling*, and used with a block or convolutional code C over A, the result is a geometrically uniform code in sequence space, called a *generalized coset code*. Most known good trellis codes are based on convolutional codes over the binary group $\mathbf{Z}_2$ (or equivalently the binary field GF(2)) and on geometrically uniform partitions that admit binary isometric labelings. Such trellis codes are geometrically uniform; consequently,

- the sets of distances (distance profile) from any trellis code sequence to all other code sequences are all the same;
- all decision regions (Voronoi regions) have the same shape.

The last property guarantees that the error probability over a Gaussian channel does not depend on which trellis code sequence is transmitted, and is also important for shaping [14] and quantization applications.

In [11], it was shown that a number of good lattice-type trellis codes, including the four-state two-dimensional code of Ungerboeck [12] and most of the Wei codes [15], are translates of linear codes, which implies they are geometrically uniform. Other good codes, including the remaining Ungerboeck codes and most of the Calderbank–Sloane [13] codes, are not translates of linear codes; however, almost all of these codes are generalized coset codes, and therefore are geometrically uniform as well.

In Section V, we discuss other notions of uniformity. In his original paper, Ungerboeck [12] showed that if his codes were used with sufficiently symmetrical signal sets, then the minimum Euclidean distance between code sequences could be determined exactly by evaluating the minimum weight over all binary error events, with an appropriately defined weighting function. Zehavi and Wolf [16] have considered similar symmetry conditions on weight profiles that permit distances, multiplicities and performance bounds to be calculated from a state diagram having no more states than the code itself. We show that all geometrically uniform codes have Ungerboeck–Zehavi–Wolf (UZW) symmetry.

Calderbank and Sloane [13] introduced a property called *regularity* that is sufficient to ensure that a coset code is distance-invariant, so that important parameters such as the minimum squared distance and multiplicity of the code do not depend on the code sequence actually transmitted. Benedetto *et al.* [17] independently introduced a similar property, called *superlinearity*. We show that geometrical uniformity implies regularity.

Section VI discusses the computation of distance distributions of geometrically uniform codes, using generating function techniques. Section VII suggests a number of lines of further research.

II. Isometries and Geometric Uniformity

We begin by defining isometries and giving a few examples. We discuss isometry groups in Euclidean N-space, and certain subgroups. Finally, we define geometric uniformity, and state important properties of geometrically uniform signal sets. For a good recent undergraduate text that covers the necessary background in group theory, see Armstrong [18].

A. Introduction to Isometries

An isometry is a distance-preserving transformation of some metric space. The usual example is Euclidean N-space, which is real N-space $\boldsymbol{R}^N$ with the Euclidean distance metric, in which the squared distance is

$$d^2(x,y) = \|x-y\|^2, \qquad x,y \in \boldsymbol{R}^N.$$

Here the squared norm $\|x\|^2$ is the inner product $(x,x) = x^T x$, where the vector x is regarded as a column vector, and the superscript T signifies the transpose.

Definition 1: An *isometry* u of Euclidean N-space $\boldsymbol{R}^N$ is a transformation u: $\boldsymbol{R}^N \to \boldsymbol{R}^N$ that preserves Euclidean distances,

$$\|u(x)-u(y)\|^2 = \|x-y\|^2, \qquad x,y \in \boldsymbol{R}^N,$$

where $u(x)$ and $u(y)$ are the images of x and y under the transformation u.

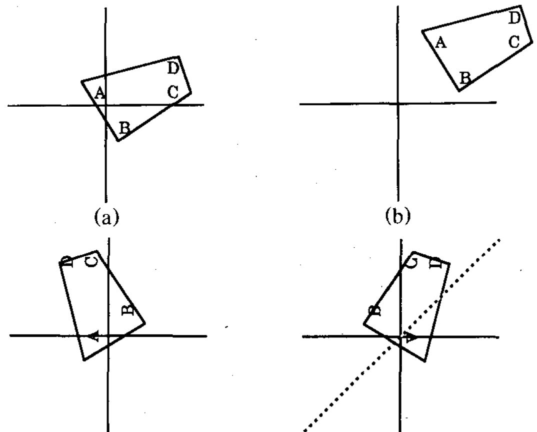

Fig. 1. Effect of various isometries of $\boldsymbol{R}^2$ on a figure ABCD. (a) Original. (b) Translation. (c) Pure rotation. (d) Pure reflection.

Important classes of isometries are:

1) *Translations*: $t_\tau(x) = x + \tau$, for any $\tau \in \boldsymbol{R}^N$.
2) *Orthogonal transformations*: $r_R(x) = Rx$, where R is an *orthogonal matrix*; i.e., R is an $N \times N$ matrix such that $R^T R = I$, the identity matrix. It follows that $\det R = \pm 1$. If $\det R = +1$, then r_R is said to be a *pure rotation*, or proper rotation; if $\det R = -1$, then r_R is said to be a rotation with reflection, or *improper rotation*; if $\det R = -1$ and $R^2 = I$ (R has order 2), then r_R is said to be a *pure reflection*.

Fig. 1 illustrates a translation, a pure rotation, and a pure reflection in $\boldsymbol{R}^2$.

Since $R^T R = I$, the columns $\{r_j, 1 \le j \le N\}$ of an orthogonal matrix R are a set of N orthonormal N-tuples, since the inner products must satisfy $(r_j, r_{j'}) = r_j^T r_{j'} = \delta_{j-j'}$. Thus the columns may be taken as an orthonormal basis of N-space $\boldsymbol{R}^N$, and an orthogonal transformation $r_R(x) = Rx$ amounts to a change of basis of $\boldsymbol{R}^N$.

It is shown in geometry that any isometry of $\boldsymbol{R}^N$ can be uniquely expressed as the composition of an orthogonal transformation r_R with a translation t_τ; i.e., as an affine transformation $u_{R,\tau}$ defined by

$$u_{R,\tau}(x) = Rx + \tau,$$

where R is an orthogonal matrix, $R^T R = I$, and τ is an arbitrary element of $\boldsymbol{R}^N$. The orthogonal transformation r_R, or its matrix R, is sometimes said to be the *linear constituent* of $u_{R,\tau}$; the translation t_τ, or its vector τ, is then the *translation constituent* of $u_{R,\tau}$.

B. Congruence and Similarity

Geometrical equivalence is defined by isometries. (Indeed, geometry has been said to be the study of properties of figures that are invariant under isometries.)

Definition 2: A *geometrical figure* S is any set of points of Euclidean N-space $\boldsymbol{R}^N$. The *image* of a figure S under an isometry u is denoted as $u(S)$. Two figures S_1 and S_2 are *geometrically congruent* (or simply *congruent*) if there exists an isometry u such that $u(S_1) = S_2$. If S_1 and S_2 are geometrically congruent, then we say that S_1 and S_2 *have the same shape*.

A broader notion of geometrical equivalence is similarity, defined as follows.

Definition 2′: Two figures S_1 and S_2 are *geometrically similar* (or simply *similar*) if there exists an isometry u and a scalar $\alpha > 0$ such that $u(\alpha S_1) = S_2$. If S_1 and S_2 are geometrically similar, then we say that S_1 is a *version* of S_2, and vice versa.

C. Symmetry Groups

The symmetries of a figure S are characterized by isometries.

Definition 3: An isometry u that leaves S invariant, $u(S) = S$, is a *symmetry* of S. The symmetries of S form a group under composition, the *symmetry group* $\Gamma(S)$ of S.

The set of all translation symmetries of S is a subgroup of $\Gamma(S)$, called the *translation symmetry group* $T(S)$ of S. The set of translation symmetry vectors $\Lambda(S) = \{\tau: t_\tau \in T(S)\}$ is an N'-dimensional lattice, $N' \le N$, called the *lattice of* S. The lattice $\Lambda(S)$ is a free, abelian group of *rank* N':

$$\Lambda(S) = \{Gx: x \in \mathbf{Z}^{N'}\},$$

where G is a generator matrix of N' linearly independent vectors of $\Lambda(S)$, and x runs through all integer N'-tuples. The translation symmetry group $T(S)$ is also the translation symmetry group $T[\Lambda(S)]$ of the lattice $\Lambda(S)$:

$$T(S) = T[\Lambda(S)] = \{t_\lambda, \lambda \in \Lambda(S)\}.$$

The symmetry group $\Gamma(S)$ is a group of orthogonal transformations if and only if the rank of $\Lambda(S)$ is 0, so that the translation symmetry group is the trivial group $\{0\}$. The lattice $\Lambda(S)$ spans $\boldsymbol{R}^N$ (its generator matrix G is a basis for $\boldsymbol{R}^N$) if and only if $\Lambda(S)$ has full rank N.

The translation symmetry group $T(S)$ is a normal subgroup of $\Gamma(S)$. The quotient group $\Gamma(S)/T(S)$ is the *linear constituent group* $\Pi(S)$. The elements of $\Pi(S)$ are the cosets of $T(S)$ in $\Gamma(S)$; each coset may be characterized as the set of all symmetries of S that have the same linear constituent R. (Such a coset need not include the orthogonal transformation $u_{R,0}$: $x \to Rx$.) The linear constituent group $\Pi(S)$ is thus isomorphic to the group $\tilde{\Pi}(S)$ of $N \times N$ orthogonal matrices R that appear as linear constituents of the elements of $\Gamma(S)$.

The set of all pure (proper) rotations in $\tilde{\Pi}(S)$ ($\det R = +1$) is a subgroup of $\tilde{\Pi}(S)$, called the *pure rotation group* $\Pi'(S)$ of S. The group $\Pi'(S)$ is either equal to all of $\tilde{\Pi}(S)$ (if there are no linear constituents with $\det R = -1$), or else is a proper normal subgroup of index 2 in $\tilde{\Pi}(S)$. In the latter case, $\tilde{\Pi}(S)$ is the union of $\Pi'(S)$ and its coset $R_- \cdot \Pi'(S)$, where R_- is any element of $\tilde{\Pi}(S)$ that has determinant -1. If there is a pure reflection R_- in $\tilde{\Pi}(S)$, then $V = \{I, R_-\}$ is a two-element subgroup of $\tilde{\Pi}(S)$, and we write $\tilde{\Pi}(S) = V \cdot \Pi'(S)$, meaning that every element of the orthogonal matrix group $\tilde{\Pi}(S)$ may be written uniquely as the composition of an element of the pure reflection group V with an element of the pure rotation group $\Pi'(S)$.

D. Geometrically Uniform Signal Sets

For digital communications, the geometrical figures of interest are sets S of discrete points (i.e., if $s \in S$, then there is a neighborhood $N(s)$ of s such that $N(s) \cap S = s$). Such figures will be called *signal sets*.

Geometrically uniform signal sets may then be defined as follows.

Definition 4: A signal set S is *geometrically uniform* if, given any two points s and s' in S, there exists an isometry $u_{s,s'}$ that transforms s to s' while leaving S invariant:

$$u_{s,s'}(s) = s',$$
$$u_{s,s'}(S) = S.$$

In other words, S is geometrically uniform if the action of its symmetry group $\Gamma(S)$ on S is *transitive*; or, if the *orbit* of any point $s_0 \in S$ under $\Gamma(S)$ is S,

$$S = \{u(s_0),\ u \in \Gamma(S)\}.$$

A finite geometrically uniform signal set S—i.e., a Slepian-type group code—will be called a *uniform constellation*; an infinite geometrically uniform signal set will be called a *regular array*.

In general, the symmetry group $\Gamma(S)$ of a geometrically uniform signal set S is larger than necessary to generate S. For example, there are eight symmetries of the square (the four-point signal set $S = \{\pm 1, \pm 1\}$); these symmetries are the elements of the *dihedral group* D_4, which may be represented by the following group of 2×2 orthogonal matrices:

$$\begin{matrix}1 & 0\\ 0 & 1,\end{matrix}\quad \begin{matrix}0 & -1\\ 1 & 0,\end{matrix}\quad \begin{matrix}-1 & 0\\ 0 & -1,\end{matrix}\quad \begin{matrix}0 & 1\\ -1 & 0,\end{matrix}\quad \begin{matrix}1 & 0\\ 0 & -1,\end{matrix}\quad \begin{matrix}0 & 1\\ 1 & 0,\end{matrix}\quad \begin{matrix}-1 & 0\\ 0 & 1,\end{matrix}\quad \begin{matrix}0 & -1\\ -1 & 0.\end{matrix}$$

The pure rotation group R_4 is the set of rotations by integer multiples of 90°, and corresponds to the first four matrices. The group R_4 is isomorphic to the group $\mathbf{Z}_4$ of integers modulo 4, and suffices to generate all points of the square S from any single point. Alternatively, the group V^2 of reflections about either axis, corresponding to the four diagonal matrices in this list, also suffices to generate all points of S from any single point; V^2 is isomorphic to $(\mathbf{Z}_2)^2$.

Definition 5: A *generating group* $U(S)$ of S is a subgroup of the symmetry group $\Gamma(S)$ that is minimally sufficient to generate S from an arbitrary initial point s_0 of S. That is, if $U(S)$ is a generating group of S, and $s_0 \in S$, then S is the orbit of s_0 under $U(S)$, $S = \{u(s_0),\ u \in U(S)\}$, and the map m: $U(S) \to S$ defined by $m(u) = u(s_0)$ is one-to-one.

Such a map from $U(S)$ to S therefore induces a group structure on S that is isomorphic to the generating group $U(S)$.

A geometrically uniform signal set S need not have a generating group. There is a famous counterexample due to Slepian [19] in which S consists of 10 points in five dimensions. We consider only signal sets S that have a generating group. (More generally, if S is geometrically uniform, and $\mathrm{Stab}(s_0)$ is the subgroup of $\Gamma(S)$ that fixes s_0, then the map m: $\Gamma(S) \to S$ defined by $m(u) = u(s_0)$ maps the left cosets of $\mathrm{Stab}(s_0)$ to the points of S, and this map can be used in a more general development. For example, as we have seen, in Slepian's counterexample, $\Gamma(S)$ is a group of order 120, and $|\mathrm{Stab}(s_0)| = 12$. If $\mathrm{Stab}(s_0)$ is nontrivial, then it cannot be a normal subgroup of $\Gamma(S)$ [20].)

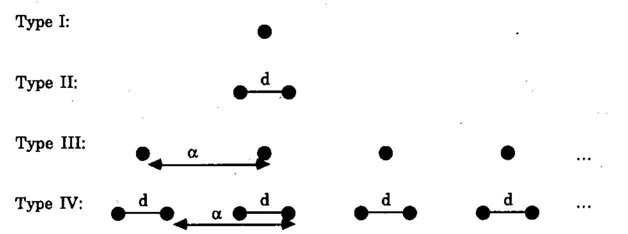

Fig. 2. Four types of geometrically uniform signal sets in one dimension.

Highly symmetrical signal sets S often have multiple possible generating groups $U(S)$. For example, as we have seen, the four-point set $S = \{\pm 1, \pm 1\}$ has two possible generating groups, and the two groups are not isomorphic.

E. One-Dimensional Geometrically Uniform Signal Sets

It is easy and instructive to characterize all geometrically uniform signal sets in one dimension. There are exactly four types, illustrated in Fig. 2. Types I and II are one- and two-point uniform constellations. There cannot be more than two points in a uniform constellation; the constellation cannot "look the same" from an end point as from an interior point, since the latter has two nearest neighbors, while the former has only one. Types III and IV are regular arrays. Since every point in a one-dimensional regular array has two nearest neighbors, and since the set of two nearest-neighbor distances must be the same for every point, these are the only two possibilities.

Any single point (Type I) is trivially a uniform constellation. Its symmetry group $\Gamma(S)$ has two elements, the identity and a reflection about the point. Its generating group $U(S)$ must be the single-element group consisting of the identity.

Any set S of two points (Type II) is a uniform constellation. Its unique generating group $U(S) = \Gamma(S)$ is the two-element group V consisting of the identity and a reflection about the midpoint, which interchanges the points. Any two-point set S is characterized by a single parameter d, up to geometrical congruence, as illustrated in Fig. 2.

A Type III regular array is geometrically similar to the set $\mathbf{Z}$ of integers. In other words, a Type III array S is a version $S = \alpha\mathbf{Z} + a$ of $\mathbf{Z}$. A natural generating group $U(S)$ for S is $T(\alpha\mathbf{Z}) = \{t_\lambda,\ \lambda \in \alpha\mathbf{Z}\}$, the set of all translations t_λ by elements λ of $\alpha\mathbf{Z}$; i.e., the translation symmetry group of S. A one-dimensional translation symmetry group $T(\alpha\mathbf{Z})$ is isomorphic to $\mathbf{Z}$. Thus a Type III array is characterized by a single parameter α, up to geometrical congruence, as illustrated in Fig. 2.

The symmetry group $\Gamma(S)$ of a Type III array $S = \alpha\mathbf{Z} + a$ includes, in addition to $T(\alpha\mathbf{Z})$, reflections about every point and every midpoint. Thus $\Gamma(S) = V \cdot T(\alpha\mathbf{Z})$, where V is a two-element group consisting of the identity and a single reflection. Such a symmetry group $\Gamma(S)$ is isomorphic to the *infinite dihedral group* D_∞. The translation symmetry group $T(\alpha\mathbf{Z})$ is a proper subgroup of the symmetry group $\Gamma(S)$, with index 2 in $\Gamma(S)$. The linear constituent group V is a two-element subgroup that is isomorphic to $\mathbf{Z}_2$. ($\Gamma(S)$ is a semidirect product of $T(\alpha\mathbf{Z})$ by V.)

An alternative generating group for a Type III array is the subgroup $U(S) = V \cdot T(2\alpha\mathbf{Z})$ of $\Gamma(S)$, since S may be regarded as the union of even and odd points, and a reflection about a midpoint interchanges even and odd points. Of course the index of $V \cdot T(2\alpha\mathbf{Z})$ in $\Gamma(S)$ is also 2. We shall see below that the group $V \cdot T(2\alpha\mathbf{Z})$, which includes reflections and is isomorphic to D_∞, is actually more useful for coding.

Given a general initial point s_0, the symmetry group $\Gamma(S) = V \cdot T(\alpha\mathbf{Z})$ generates an array of Type IV, which is not a version of $\mathbf{Z}$; the additional symmetry of the Type III array occurs only with special initial points. Such an array S may be regarded as a Type III array with every point split into two points; i.e., S is the union of two translates $\alpha\mathbf{Z} + a_1$ and $\alpha\mathbf{Z} + a_2$ of a single scaled integer lattice $\alpha\mathbf{Z}$. Its unique generating group $U(S)$ is its symmetry group $\Gamma(S)$, which is the infinite dihedral group $V \cdot T(\alpha\mathbf{Z})$. A Type IV array is characterized by two parameters (α, d), up to geometrical congruence (where $d = |a_1 - a_2|$), as illustrated in Fig. 2.

It is interesting to note that the asymmetrical signal sets used by Calderbank and Mazo [21] are in fact finite subsets of geometrically uniform Type IV arrays.

All four types are special cases of the Type IV array. The Type III array is the special case in which $d = 0$; the Type II constellation is the special case in which $\alpha = 0$; and the Type I constellation is the special case in which $\alpha = d = 0$. Alternatively, Types III, II, and I are obtained as the orbits of an arbitrary initial point under subgroups of the symmetry group $V \cdot T(\alpha\mathbf{Z})$ of the type IV array: $T(\alpha\mathbf{Z})$, V, and the trivial identity group, respectively.

F. Geometrically Uniform Signal Sets in Communications

Most signal sets S used in digital communications are geometrically uniform.

Example 1: MPSK signal constellations $S = \{\omega^j s_0, 0 \le j \le M-1\}$, where ω is a primitive complex Mth root of unity, and s_0 is any nonzero complex number. A natural generating group of S is $U(S) = R_M$, the set of rotations by multiples of 360°/M. This group is isomorphic to $\mathbf{Z}_M$, the additive group of integers modulo M. The symmetry group of S is $\Gamma(S) = V \cdot R_M$, the set of all compositions of elements of R_M with elements of a two-element reflection group V consisting of the identity and a reflection about the line between any point or midpoint and the origin. This group is isomorphic to the *M-ary dihedral group* D_M. (Here $\Gamma(S)$ is a semidirect product of R_M by V.)

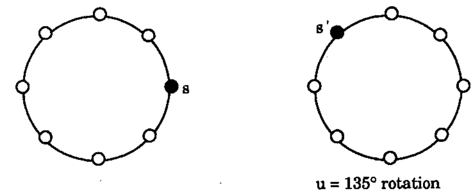

Fig. 3. An 8PSK signal constellation is a uniform constellation.

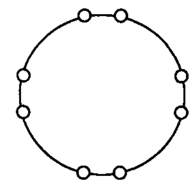

Fig. 4. Second type of two-dimensional uniform constellation.

For example, Fig. 3 shows an 8PSK signal constellation. The constellation is invariant under the group $U(S) = R_8$ of rotations by multiples of 45°, and any point $s \in S$ can be transformed into any other point $s' \in S$ by an element of R_8, which is isomorphic to $\mathbf{Z}_8$.

The symmetry group $\Gamma(S) = V \cdot R_M$ also includes reflections about the lines joining points or midpoints to the origin. If M is even, then S has an alternative generating group $U(S)$ of the type $V \cdot R_{M/2}$, which is isomorphic to the dihedral group $D_{M/2}$. This is the symmetry group of the only other type of two-dimensional uniform constellation, illustrated in Fig. 4. This type of uniform constellation may be obtained from an $(M/2)$-PSK constellation by splitting points, or as the orbit of a general initial point under $V \cdot R_{M/2}$, in complete analogy to the Type IV array of the previous section. (Such constellations have been proposed for use with trellis codes by Divsalar *et al.* [22], [23].)

As another example, the four vertices of a square such as $S = \{\pm 1, \pm 1\}$ may be regarded as a 4PSK constellation; its symmetry group is $V \cdot R_4$, and two alternative generating groups are R_4 and $V \cdot R_2 \simeq V^2$, as discussed in Section II-D.

In general, any N-dimensional uniform constellation must have a trivial translation symmetry group, and must therefore be a Slepian-type group code (see the Introduction), consisting of a finite set of points on the surface of an N-sphere. The restriction to spherical signal sets is quite limiting in general, although required in certain applications.

Example 2: Binary signal constellations $S = \{\pm d\}$. As discussed in Section II-E, any two-point set is a uniform constellation. In one dimension, the unique generating group of $S = \{\pm d\}$ is the symmetry group $\Gamma(S) = V$, con-

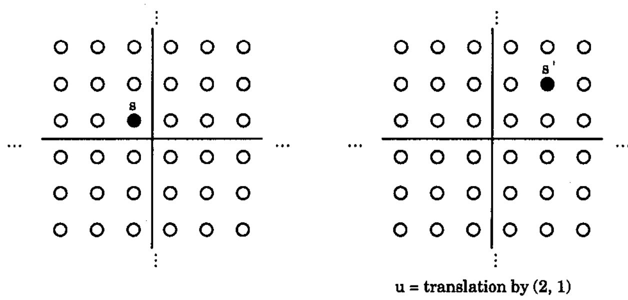

Fig. 5. The half-integer grid is a regular array.

sisting of the identity ($R=[1]$) and a reflection about the origin ($R=[-1]$); $\Gamma(S)$ is isomorphic to Z_2.

More generally, an N-dimensional *hypercube* constellation $S=\{(\pm d,\cdots,\pm d)\}$ is a uniform constellation. This follows from the fact that the Cartesian product of geometrically uniform signal sets is geometrically uniform, which can be shown by considering the composition of isometries of each component individually. A natural generating $U(S)$ is the 2^N-element axial reflection group V^N, the N-fold Cartesian product of the reflection group V acting on each of the N coordinates individually, which is represented by the set of diagonal orthogonal matrices $A=\mathrm{diag}(\pm 1,\cdots,\pm 1)$. The group V^N is isomorphic to $(Z_2)^N$. Again, the case $N=2$ corresponds to the example of Section II-D as well as to 4PSK. For $N\geq 2$, V^N is a proper subgroup of the symmetry group of the hypercube, as we saw for $N=2$.

Example 3: Lattice-type signal sets $S=\Lambda+\tau$. For example, the *half-integer grid* $S=\boldsymbol{Z}^2+(1/2,1/2)$, a translate of the two-dimensional integer lattice $\boldsymbol{Z}^2$, is illustrated in Fig. 5. One generating group $U(S)$ is the translation isometry group $T(\Lambda)$ of Λ, which in this case is $T(\boldsymbol{Z}^2)$, the set of translations by integers in each coordinate.

The symmetry group $\Gamma(S)$ of the half-integer grid includes not only the translation symmetry group $T(\boldsymbol{Z}^2)$ but also the eight-element dihedral group D_4 of symmetries of the square, which is its linear constituent group. Indeed, $\Gamma(S)$ is the set of all compositions of elements of D_4 with elements of $T(\boldsymbol{Z}^2)$, and may be written as $\Gamma(S)=D_4\cdot T(\boldsymbol{Z}^2)$. (In this case $\Gamma(S)$ is a semidirect product of $T(\boldsymbol{Z}^2)$ by D_4.)

Like the square, the half-integer grid has more than one generating group. In addition to $T(\boldsymbol{Z}^2)$, the groups $R_4\cdot T(2\boldsymbol{Z}^2)$ or $V^2\cdot T(2\boldsymbol{Z}^2)$ are alternative generating groups, since an initial point such as $s_0=(1/2,1/2)$ may be moved to the vertices of a square by elements of R_4 or V^2, and then translations of the four vertices by elements of $2\boldsymbol{Z}^2$ can generate all points in S. These three possible generating groups are all subgroups of $\Gamma(S)$ of index 8. The group $V^2\cdot T(2\boldsymbol{Z}^2)$ is the Cartesian product of the one-dimensional group $V\cdot T(2\boldsymbol{Z})$ with itself, which again will turn out to be the best choice for coding. The group $R_4\cdot T(2\boldsymbol{Z}^2)$ may be of interest in connection with four-way rotational invariance.

G. Regular Arrays as Crystals

It is conceptually intriguing to think of regular arrays as crystals. (No results from crystallography will actually be used in this paper.)

In crystallography [9], a *crystal structure* is defined as a discrete array S in Euclidean N-space $\boldsymbol{R}^N$ whose translation symmetry group $T(S)$ has rank N. The N-dimensional translation symmetry lattice $\Lambda(S)$ is called the *lattice of the crystal*. A crystal structure S is thus necessarily infinite, and its elements span $\boldsymbol{R}^N$.

A regular array S with a translation symmetry group of rank N is therefore a crystal structure; a crystal structure with a transitive symmetry group is a regular array.

A *crystallographic symmetry operation* (CSO) is an isometry of $\boldsymbol{R}^N$ that leaves a crystal structure S invariant; i.e., a symmetry of S. The linear constituent of a CSO is representable by an invertible (*unimodular*) integer matrix.

The *space group* of a crystal structure S is its symmetry group $\Gamma(S)$. A *crystallographic group* is any subgroup of a space group. The translation symmetry group $T(S)$ is a normal, free, abelian rank-N subgroup of the space group. The *point group* of S is the linear constituent group $\Gamma(S)/T(S)$. The point group is always finite, and may be represented by a finite group of unimodular $N\times N$ matrices.

This leads to the famous *crystallographic restriction*: there exists an N-dimensional linear constituent R of order n ($R^n=I$), if and only if $\varphi(n)\leq N$, where $\varphi(n)$ is the Euler function of n (the number of integers less than n that are relatively prime to n) [10]. Thus in one dimension, R can have order only 1 or 2 ($R=[1]$ or $[-1]$). In two or three dimensions, R can have order 1, 2, 3, 4, or 6; the half-integer grid of Fig. 5 has symmetries of orders 2 and 4, while the hexagonal lattice A_2 has symmetries of orders 3 and 6. (There is a translate of A_2, which has a symmetry of order 3 with no fixed points, but there exists no such translate with a symmetry of order 6.) In four or five dimensions, R can have order 1, 2, 3, 4, 5, 6, 8, 10, or 12; and so forth.

The classification of all three-dimensional space groups was completed late in the nineteenth century. In 1900 Hilbert posed the question (Hilbert's 18th problem) whether the number of classes of space groups in every dimension N is finite. The question was answered in the affirmative in the following decade by Bieberbach. The number of such classes, in one classification, is 2 in one dimension ($\simeq \boldsymbol{Z}$ and D_∞), and 17, 219, and 4783 for $N=2$, 3, and 4, the last having been established by computer search [9]. For coding purposes, it seems that the most useful space groups are those with the largest point groups.

H. Symmetry Properties of Geometrically Uniform Signal Sets

A geometrically uniform signal set S has the important property of looking the same from any to its points. This property implies that any arbitrary point of S may be taken as the "center of the universe," and that all geometric properties relative to that point do not depend on which point is chosen.

In particular, two important geometric properties are the following.

Definition 6: The *Voronoi region* $\mathbb{R}_V(s)$ associated with any point $s \in S$ is the set of all points in $\boldsymbol{R}^N$ that are at least as close to s as to any other point $s' \in S$:

$$\mathbb{R}_V(s) = \{x \in \boldsymbol{R}^N: \|x - s\|^2 = \min_{s' \in S} \|x - s'\|^2\}.$$

The *global distance profile* DP(s) associated with any point $s \in S$ is the set of distances to all points in S:

$$\mathrm{DP}(s) = \{\|s - s'\|^2, s' \in S\}.$$

Theorem 1 (Geometrical Uniformity): If S is a geometrically uniform signal set, then a) all Voronoi regions $\mathbb{R}_V(s)$ have the same shape, and indeed $\mathbb{R}_V(s') = u_{s,s'}[\mathbb{R}_V(s)]$, where $u_{s,s'}$ is any isometry that takes s to s'; b) the global distance profile DP(s) is the same for all $s \in S$.

Proof: By definition, $x \in \boldsymbol{R}^N$ is in $\mathbb{R}_V(s)$ if and only if

$$\|x - s\|^2 = \min_{s'' \in S} \|x - s''\|^2.$$

If $s' \neq s$ is another point in S, then by the definition of geometrical uniformity there is an isometry $u_{s,s'}$ such that $u_{s,s'}(s) = s'$, $u_{s,s'}(S) = S$. If $u_{s,s'}$ is any such isometry, then $u_{s,s'}(x)$ is in $\mathbb{R}_V(s')$, because

$$\begin{aligned}\|u_{s,s'}(x) - s'\|^2 &= \|u_{s,s'}(x) - u_{s,s'}(s)\|^2\\ &= \|x - s\|^2\\ &= \min_{s'' \in S} \|x - s''\|^2\\ &= \min_{s'' \in S} \|u_{s,s'}(x) - u_{s,s'}(s'')\|^2\\ &= \min_{y \in S} \|u_{s,s'}(x) - y\|^2,\end{aligned}$$

where we observe that as s'' ranges through S, so does $y = u_{s,s'}(s'')$. Similarly,

$$\begin{aligned}\mathrm{DP}(s) &= \{\|s - s''\|^2, s'' \in S\}\\ &= \{\|u_{s,s'}(s) - u_{s,s'}(s'')\|^2, s'' \in S\}\\ &= \{\|s' - u_{s,s'}(s'')\|^2, s'' \in S\}\\ &= \{\|s' - y\|^2, y \in S\}\\ &= \mathrm{DP}(s').\end{aligned}$$

□

The fact that the shape of every Voronoi region of S is the same is a very strong symmetry property, because the shape of the Voronoi region determines almost all properties of S that are important for communications. We may take any Voronoi region $\mathbb{R}_V(s)$ as a typical Voronoi region, and call it *the Voronoi region* $\mathbb{R}_V(S)$ of S. From any such region $\mathbb{R}_V(s)$, we can determine the following properties:

a) the *fundamental volume* $V(S)$ of S, namely the volume of $\mathbb{R}_V(S)$, which is finite if and only if the lattice of S has full rank N; or equivalently the *density* $D(S) = 1/V(S)$ of S (the number of points of S per unit volume);

b) the *local distance profile* $\mathrm{DP}_V(S) = \{\|s - s'\|^2, s' \in S_V\}$, where S_V is the finite set of *relevant* or *face-defining* near neighbors s' to s, one corresponding to each face of the convex polytope $\mathbb{R}_V(s)$; a tighter union bound on error probability can be obtained from $\mathrm{DP}_V(S)$ than the usual union bound based on the global distance profile DP(S);

c) in particular, $\mathrm{DP}_V(S)$ determines the *minimum squared distance* $d^2_{\min}(S)$ between points in S, or equivalently the *packing radius* $d_{\min}(S)/2$, and the *multiplicity* (kissing number) $K(S)$, the number of such nearest neighbors;

d) the *second moment* $E_u[\|x - s\|^2]$ of $\mathbb{R}_V(s)$ about s, under a uniform distribution $p_u(x) = 1/V(S)$ over $\mathbb{R}_V(s)$, which is of importance for quantization and for shaping;

e) the squared *covering radius* $r^2_{\max} = \max\{\|x - s\|^2, x \in \mathbb{R}_V(s)\}$, the number of *deep holes* at this radius, and indeed the distances to all vertices of $\mathbb{R}_V(s)$;

f) the *error probability*

$$\Pr(E) = 1 - \int_{\mathbb{R}_V(s)} (2\pi\sigma^2)^{-N/2} \mathrm{e}-\left(\|z - s\|^2/2\sigma^2\right) dz$$

when the point s is transmitted over a memoryless Gaussian channel with noise variance σ^2 per dimension; or, more generally, the error probability over any channel in which the probability density $p(z|s)$ of the received signal z depends only on the squared distance $\|z - s\|^2$, and decreases with squared distance.

Theorem 1 asserts that if S is geometrically uniform, then all of these properties are the same for every point in S. Therefore, without loss of generality, the properties of S may be considered from the perspective of any point $s \in S$.

For communication at finite data rates, one may use a finite constellation $C(S, \mathbb{R})$ consisting of the points in an infinite regular array S that lie within a bounding region $\mathbb{R}$ of finite volume. For example, a standard 4×4 square QAM constellation is not a uniform constellation, but it can be described as the set of points in the regular array S of Fig. 5 (the half-integer grid) that lie within a square $\mathbb{R}$ of side 4. In this case the Voronoi regions of the points in the interior of $C(S, \mathbb{R})$ will still have the same shape, but the Voronoi regions of the outer points will be modified, due to boundary effects. One often gets good approximations by ignoring such effects, particularly when $C(S, \mathbb{R})$ is large. In other words, one may say that "$C(S, \mathbb{R})$ is geometrically uniform, up to boundary effects."

Fig. 2 illustrates a one-dimensional regular array that is not simply the translate of a lattice. In two dimensions, there are many more types of regular arrays (37, in the classification scheme of [24, pp. 240–243]). Fig. 6 illus-

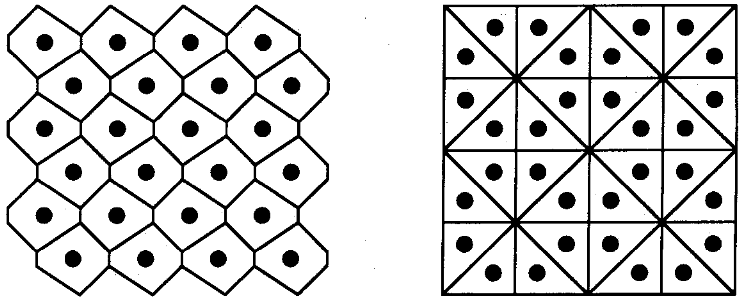

Fig. 6. Two regular arrays in $\boldsymbol{R}^2$ that are not simply translates of lattices.

trates two such two-dimensional regular arrays (or crystal structures), with their Voronoi regions.

III. Partitions and Labelings of Geometrically Uniform Signal Sets

A signal space code is defined by a partition of a signal set into subsets, a labeling of the subsets, and a label code that specifies a sequence of subsets via a label sequence. These key ideas were called "mapping by set partitioning" by Ungerboeck [12].

This section shows how a normal subgroup U' of a generating group $U(S)$ induces a partition of a geometrically uniform signal set S into geometrically uniform subsets. The natural labeling of such a partition is by a label group that is isomorphic to the quotient group $U(S)/U'$. Most practical label codes are binary; therefore we focus particularly on cases in which $U(S)/U'$ is isomorphic to a group $(\mathbf{Z}_2)^n$ of binary n-tuples for some integer n.

A. Geometrically Uniform Partitions

From elementary group theory, a subgroup G' of a group G induces a partition of G into *cosets* of G' in G. The left and right cosets of G' in G are identical and form a well-defined *quotient group* G/G' if and only if G' is a *normal* (self-conjugate) subgroup of G. The *index* of G' in G is the order $|G/G'|$ of the quotient group G/G'.

Let U' be a normal subgroup of a generating group $U(S)$ of a geometrically uniform signal set S, and let $U(S)/U'$ be the corresponding quotient group of cosets of U' in $U(S)$. If S is the orbit of a point s_0 under $U(S)$, then the orbits of s_0 under the cosets of U' are disjoint subsets of S whose union is S. The orbit of s_0 under U' itself will be denoted by S', and the partition of S induced by U' will be denoted by S/S'.

Definition 6: A *geometrically uniform partition* S/S' is a partition of a geometrically uniform signal set S with a generating group $U(S)$ that is induced by a normal subgroup U' of $U(S)$. The elements of the partition S/S' are the subsets of S that correspond to the cosets of U' in $U(S)$.

(In crystallography, partitions are studied under the rubric of *color symmetry*; the subsets of S are regarded as having different colors.)

Theorem 2 (Geometrically Uniform Partitions): Let S/S' be a geometrically uniform partition. Then the subsets of S in this partition are geometrically uniform, mutually congruent, and have U' as a common generating group.

Proof: If U' is a normal subgroup of $U(S)$, then the left and right cosets of U' are identical. A left coset of U' is the set u_aU' of all isometries that are the composition of some fixed isometry u_a with all $u \in U'$, and the corresponding subset $S'(a)$ is

$$S'(a) = \bigcup_{u \in U'} u_a[u(s_0)] = u_a\Big[\bigcup_{u \in U'} u(s_0)\Big] = u_a(S').$$

Therefore every subset $S'(a)$ is congruent to S'. A right coset of U' is the set $U'u_a$ of all isometries that are the composition of some $u \in U'$ with a fixed isometry u_a, and the corresponding subset $S'(a)$ is

$$S'(a) = \bigcup_{u \in U'} u[u_a(s_0)].$$

Thus every subset $S'(a)$ is the orbit of a point $u_a(s_0)$ under all isometries in the group U', so $S'(a)$ is geometrically uniform and has generating group U'. □

In view of Theorem 2, the subgroup U' will henceforth be denoted by $U(S')$, since it is a generating group of S'.

Theorem 2 is a powerful tool for deriving geometrically uniform signal sets from a single geometrically uniform parent set, particularly if the parent set has a rich group structure. For example, the Type IV one-dimensional set, with generating group $V \cdot T(\alpha) \simeq D_\infty$, is the parent of all one-dimensional types; the Type III set is generated by the subgroup $T(\alpha\mathbf{Z}) \simeq \mathbf{Z}$, the Type II set by the subgroup $V \simeq \mathbf{Z}_2$, and the Type I set by the trivial subgroup consisting only of the identity.

If M' is any divisor of M, an MPSK signal set may be partitioned into $|R_M/R_{M'}| = M/M'$ M'-PSK signal sets by the subgroup $R_{M'}$ of its rotation isometry group R_M. The quotient group $R_M/R_{M'}$ is isomorphic to $\mathbf{Z}_{M/M'}$. When M and M' are powers of 2, partitions of the type used by Ungerboeck [12] for PSK-type trellis codes are obtained.

The hypercube constellation in N dimensions has a generating group V^N that is isomorphic to $(\mathbf{Z}_2)^N$. A linear binary (N, K) code is a subgroup of $(\mathbf{Z}_2)^N$ of order 2^K. By Theorem 2, this subgroup induces a geometrically uniform partition of the hypercube into 2^{N-K} geometrically uniform, mutually congruent subsets that correspond to the cosets of the (N, K) code. Thus not only does any linear binary code map to a geometrically uniform Slepian-type group code [1], but so also does any coset of any linear binary code.

If Λ' is any sublattice of a lattice Λ, a lattice translate $S = \Lambda + a$ may be partitioned into $|\Lambda/\Lambda'|$ cosets of Λ' by the subgroup $T(\Lambda')$ of its translation isometry group $T(\Lambda)$. The quotient group $T(\Lambda)/T(\Lambda')$ is isomorphic to Λ/Λ'. This is the essence of the lattice/coset viewpoint introduced into the field of trellis codes by Calderbank and Sloane [13].

Theorem 2 may be extended to show that if $U(S)/U(S')/U(S'')/\cdots$ is a group partition chain, then there is a corresponding chain $S/S'/S''/\cdots$ of geometrically uniform partitions, in which the subsets at each level are geometrically uniform, mutually congruent, and have a common generating group.

Note: Ginzburg [26, Remark 2.1] observed that if there is a one-to-one map between a signal set S and a group G, then a chain of subgroups of G leads to a corresponding chain partition of S. Biglieri and Elia [27] applied this observation to achieve "fair partitions" of finite "generalized group alphabets," not necessarily geometrically uniform, and to prove symmetry properties for intrasubset and intersubset distances. They comment:

> "What makes [this] work is the fact that the orthogonal matrices form a group of isometries. Hence a more abstract formulation is possible, extending to non-finite groups. As pointed out by the editor [A. R. Calderbank], lattices and sublattices equipped with isometric transformations (translations) fit this more general approach... the symmetry and homogeneity properties of [the lattice/coset alphabets of Calderbank and Sloane [13]] are almost identical to those of generalized group alphabets."

I wish to acknowledge receipt of a private communication from Calderbank to the same effect, whose significance totally escaped me at the time (*c.* 1987).

B. Isometric Labelings

Let A be any group that is isomorphic to the quotient group $U(S)/U(S')$. The group A may be used as a label alphabet for the cosets of $U(S')$ in $U(S)$, where the group isomorphism determines the one-to-one map between elements of A and the elements of $U(S)/U(S')$. The group A may further be used as a label alphabet for the subsets of S in the partition S/S', using the one-to-one map between these subsets and the elements of $U(S)/U(S')$; this determines a *label map* m: $A \to S/S'$ in which a label $a \in A$ is mapped to a subset $m(a) \in (S/S')$ of S. The identity element of A maps to S', and $|S/S'| = |U(S)/U(S')| = |A|$. The group structure induced on the partition S/S' by the group structure of $U(S)/U(S')$ may thus be simply exhibited by choosing an appropriate label alphabet A.

Definition 7: A *label group* A for a geometrically uniform partition S/S' is a group that is isomorphic to the quotient group $U(S)/U(S')$, where $U(S)$ and $U(S')$ are generating groups of S and S', respectively. The isomorphism $A \simeq U(S)/U(S')$ will be called the *label isomorphism*. The one-to-one label map m: $A \to S/S'$ defined by the composition of the label isomorphism with the induced one-to-one map $U(S)/U(S') \to S/S'$ is an *isometric labeling* of the subsets of S in the partition S/S'.

In other words, the necessary and sufficient conditions for a partition S/S' to admit an isometric labeling by a group A are as follows.

Proposition 1: A partition S/S' admits an isometric labeling by a group A if a) S is geometrically uniform; b) its subsets are geometrically uniform and mutually congruent; and c) there exist isometry groups $U(S)$ and $U(S')$ such that $U(S)$ is a generating group of $S, U(S')$ is a common generating group for the subsets of S and is a normal subgroup of $U(S)$, and A is isomorphic to $U(S)/U(S')$.

In this paper A will always be an abelian (additive, commutative) group, although many of our results are more general. The label group operation will therefore be called addition, and will be denoted by $\oplus$. The identity of A will be denoted by 0; then $m(0) = S'$.

The label isomorphism means that for every $a \in A$, there is an equivalence class U_a of isometries in the quotient group $U(S)/U(S')$, such that the effect of addition by a in the label space A is the same as the effect of any isometry u_a in the class U_a in signal space. The existence of such an isometry can be taken as the defining property of an isometric labeling:

Proposition 2: A label map m: $A \to S/S'$ is an isometric labeling if and only if for all $a \in A$ there exists an isometry u_a such that for all $b \in A$,

$$m(a \oplus b) = u_a[m(b)].$$

A similar concept is introduced in Biglieri *et al.* [25, ch. 4].

C. Elementary Properties of Isometric Labelings

Given an isometric labeling m: $A \to S/S'$, there are various simple transformations of the labels that result in alternative isometric labelings for the same set of subsets.

Property 1 (Linear Transformations of the Label Space): If m: $A \to S/S'$ is an isometric labeling, and T: $A \to A$ is any group automorphism of A, then the labeling mT: $A \to S/S'$ defined by $a \to m[T(a)]$ is also isometric.

Proof: The label isomorphism $A \simeq U(S)/U(S')$ remains an isomorphism when A is transformed by the isomorphism T. □

Property 2 (Translations of the Label Space): If m: $A \to S/S'$ is an isometric labeling, then the labeling mt_b: $A \to S/S'$ defined by $a \to m(a \oplus b)$ for any $b \in A$ is isometric.

Proof: Now $0 \in A$ is mapped to $m(b)$, and $a \in A$ is mapped to $m(a \oplus b) = u_a[m(b)]$, where u_a is any isometry in the class $U_a \in U(S)/U(S')$ that corresponds to a. □

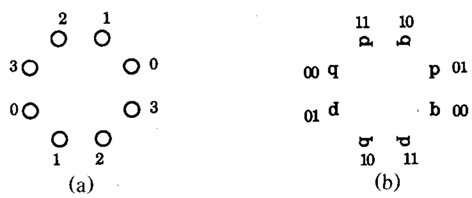

Fig. 7. Isometric labelings of four-way partitions of 8PSK signal sets.

Together, Properties 1 and 2 imply that if S/S' admits an isometric labeling with the label alphabet A, then any affine transformation $a \to T(a) \oplus b$ of A also produces an isometric labeling. It is interesting to note that an affine transformation is an isometry in label space, where the natural metric is the Hamming metric.

The following result is a useful concrete application of this principle.

Corollary 1 (Small Binary Labelings): If there exists any isometric labeling for a two-way or four-way partition S/S' with a binary label alphabet $\mathbf{Z}_2$ or $(\mathbf{Z}_2)^2$, then every possible binary labeling of S/S' is isometric.

Proof: The label spaces $\mathbf{Z}_2$ and $(\mathbf{Z}_2)^2$ can be permuted in every possible way by affine transformations. □

Finally, if any isometry u in $U(S)$ is applied to the signal space $\boldsymbol{R}^N$, then the following property holds.

Property 3 (Isometries of Signal Space): If m: $A \to S/S'$ is an isometric labeling, and u is a symmetry of S in $U(S)$, then the labeling um: $A \to S/u(S')$ defined by $a \to u[m(a)]$ is isometric.

Proof: Now $0 \in A$ is mapped into $u[m(0)] = u(S')$, and $a \in A$ is mapped into $u[m(a)] = u[u_a(S')] = u[u_a(u^{-1}[u(S')])]$, where u^{-1} is the inverse isometry to u, and u_a is any isometry in the coset $U_a \in U(S)/U(S')$ that corresponds to a. Since $U(S')$ is a normal subgroup of $\Gamma(S)$, $u^{-1}U_a u = U_a$ for all $u \in U(S)$, by the definition of normality, so this transformation simply replaces u_a by $u^{-1}u_a u$ as a representative for U_a. □

If $U(S')$ is a normal subgroup of $\Gamma(S)$, then Property 3 holds for any symmetry u in $\Gamma(S)$.

D. Examples of Isometric Labelings

An 8PSK signal set may be partitioned into four 2PSK subsets by the subgroup R_2 of its rotation isometry group R_8. The quotient group R_8/R_2 is isomorphic to $\mathbf{Z}_8/\mathbf{Z}_2 \simeq \mathbf{Z}_4$, so $\mathbf{Z}_4 = \{0,1,2,3\}$ may be used as a label alphabet A for the four subsets, as illustrated in Fig. 7(a).

Alternatively, taking the generating group as $V \cdot R_4 \simeq D_4$, the 8PSK constellation may be partitioned into four 2PSK subsets by the subgroup R_2; the quotient group $(V \cdot R_4)/R_2$ is then isomorphic to $(\mathbf{Z}_2)^2 = \{00, 01, 10, 11\}$, so $(\mathbf{Z}_2)^2$ may be used as a label alphabet A for the four subsets, as shown in Fig. 7(b). (Here we have represented the points by asymmetrical motifs to indicate the generating group.) Thus the four-way partition of the 8PSK signal set admits a binary isometric labeling by $(\mathbf{Z}_2)^2$. From the corollary of the previous section, every binary labeling of this four-way partition is thus isometric.

As another example, the half-integer grid $S = \mathbf{Z}^2 + (1/2, 1/2)$ may be partitioned into 16 subsets (cosets of $4\mathbf{Z}^2$) by the subgroup $T(4\mathbf{Z}^2)$ of its translation symmetry group $T(\mathbf{Z}^2)$. The quotient group $T(\mathbf{Z}^2)/T(4\mathbf{Z}^2)$ is isomorphic to $Z^2/4\mathbf{Z}^2 \simeq (\mathbf{Z}/4\mathbf{Z}) \times (\mathbf{Z}/4\mathbf{Z}) \simeq \mathbf{Z}_4 \times \mathbf{Z}_4$, so we may use the label alphabet $A = \mathbf{Z}_4 \times \mathbf{Z}_4$ for this 16-way partition, as shown in Fig. 8(a).

Alternatively, using the generating group $R_4 \cdot T(2\mathbf{Z}^2)$ for S, the same 16-way partition may be induced by the subgroup $T(4\mathbf{Z}^2)$; the quotient group $R_4 \cdot T(2\mathbf{Z}^2)/T(4\mathbf{Z}^2)$ is isomorphic to $\mathbf{Z}_4 \times \mathbf{Z}_2 \times \mathbf{Z}_2$, so this partition admits an isometric labeling by the label alphabet $A = \mathbf{Z}_4 \times \mathbf{Z}_2 \times \mathbf{Z}_2$, as shown in Fig. 8(b), with the asymmetrical motifs exhibiting the corresponding symmetries.

Finally, using the isometry group $V^2 \cdot T(2\mathbf{Z}^2) \simeq [V \cdot T(2\mathbf{Z})]^2 \simeq (D_\infty)^2$, and the subgroup $T(4\mathbf{Z}^2) \simeq [T(4\mathbf{Z})]^2$, we obtain a quotient group that is isomorphic to $(\mathbf{Z}_2)^4$, so this partition admits an isometric labeling by the binary label alphabet $A = (\mathbf{Z}_2)^4$, as shown in Fig. 8(c). This is a two-fold Cartesian product of a binary labeling of a four-way one-dimensional partition of $\mathbf{Z} + 1/2$ induced by $[V \cdot T(2\mathbf{Z})]/T(4\mathbf{Z})$, using the binary label binary label alphabet $A = (\mathbf{Z}_2)^2$. Again, from the corollary of the previous section, any binary labeling for this four-way one-dimensional partition must be isometric.

E. Binary Isometric Labelings

Partitions S/S' that admit isometric labelings by n-bit binary label alphabets $A = (\mathbf{Z}_2)^n$—i.e., *binary isometric labelings*—are of particular interest. From the previous examples, we see that there exist binary isometric labelings for

a) four-way partitions of the 8PSK signal set; or, more generally, four-way partitions of any MPSK signal set, where M is a multiple of four, since we may take $U(S) = V \cdot R_{M/2} \simeq D_{M/2}$, $U(S') = R_{M/4} \simeq \mathbf{Z}_{M/4}$, so $U(S)/U(S') = (V \cdot R_{M/2})/R_{M/4} \simeq (\mathbf{Z}_2)^2$;
b) 16-way partitions of the half-integer grid; or, more generally,
c) 4^N-way partitions of any N-dimensional lattice-type signal set $S = \mathbf{Z}^N + \tau$, where $\mathbf{Z}^N$ is the lattice of integer N-tuples, since we may take $U(S) = [V \cdot T(2\mathbf{Z})]^N \simeq (D_\infty)^N$, $U(S') = [T(4\mathbf{Z})]^N \simeq \mathbf{Z}^N$, so $U(S)/U(S') \simeq [(V \cdot T(2\mathbf{Z}))/T(4\mathbf{Z})]^N \simeq [(\mathbf{Z}_2)^2]^N \simeq (\mathbf{Z}_2)^{2N}$.

1) Isometric Ungerboeck Labelings: There is no significant distinction between $(\mathbf{Z}_2)^n$ as an additive group and $(\mathbf{Z}_2)^n$ as an n-dimensional vector space over the binary field $\boldsymbol{F} = \mathrm{GF}(2)$. Thus when a partition admits a binary isometric labeling, the label alphabet $A = (\mathbf{Z}_2)^n$ may as well be viewed as a vector space $\boldsymbol{F}^n$. The subgroups of $(\mathbf{Z}_2)^n$ correspond to the subspaces of $(\mathbf{Z}_2)^n$ as a vector space $\boldsymbol{F}^n$; any k-dimensional subspace of $\boldsymbol{F}^n$ is isomorphic to $(\mathbf{Z}_2)^k$.

An *Ungerboeck labeling* [11], [12] of a 2^n-way partition S/S' is defined as follows. The partition S/S' is refined into a chain of n two-way partitions $S = S_0/S_1/\cdots/S_n = S'$. The labels of the 2^n subsets are chosen from $\boldsymbol{F}^n$ so

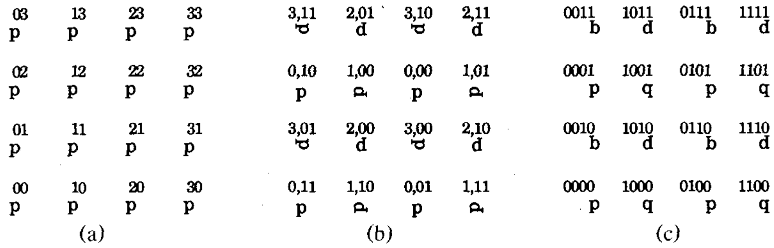

Fig. 8. Isometric labelings of 16-way partitions of the half-integer grid.

as to reflect this nested structure, in the sense that the labels of all subsets that belong to a common subset at the jth level have the same last (least significant) j bits.

There follows the *Ungerboeck distance bound*: the minimum squared distance between two distinct points with labels a and a' is at least as great as $d^2_{\min}(S_j)$, where j is the number of trailing zeroes in the mod-2 sum $a \oplus a'$, since this implies that the two points belong to a common subset at the jth level.

Theorem 4 (Ungerboeck Labelings): If S/S' is a geometrically uniform partition that admits a binary isometric labeling, and $S = S_0/S_1/\cdots/S_n = S'$ is a chain of two-way geometrically uniform partitions, then S/S' admits an isometric Ungerboeck labeling consistent with this chain.

Proof: If S/S' admits a binary isometric labeling, then there exist generating groups $U(S)$ and $U(S')$ such that $U(S)/U(S') \simeq (\mathbf{Z}_2)^n$, and if $S = S_0/S_1/\cdots/S_n = S'$ is a chain of geometrically uniform partitions, then there exists a corresponding chain of generating groups $U(S) = U(S_0)/U(S_1)/\cdots/U(S_n) = U(S')$. Let A_j stand for the set of labels included in the subset S_j of S; the corresponding chain $A = A_0/A_1/\cdots/A_n = \{0\}$ is a chain of binary vector spaces, where the dimension of A_j is $n-j$. Let the coset representative for the nonzero coset in the partition A_j/A_{j+1}, $0 \le j \le n-1$, be taken as some label $(x)^{n-j-1}10^j$, where $x \in \mathbf{Z}_2$. Then by induction $A_j/A_n = A_j$ is the $(n-j)$-dimensional binary vector space consisting of all labels $(x)^{n-j}0^j$, and the subsets at the jth level consist of the union of all nth-level subsets that have a common last j bits in their label. Thus any such labeling is an Ungerboeck labeling. □

The leading bits in an Ungerboeck labeling are called the *most significant bits*, and the trailing bits are called the *least significant bits*. This is consistent with the usage for the standard binary representation of the integers $\mathbf{Z}$, which corresponds to the lattice partition chain $\mathbf{Z}/2\mathbf{Z}/4\mathbf{Z}/\cdots$. Then the Ungerboeck distance bound can be stated in terms of the agreement of labels in their j least significant bits. Usually the chain is chosen so that the minimum squared distance $d^2_{\min}(S_j)$ between distinct points within sets at the jth level increases as rapidly as possible.

For example, the four-way partition of the 8PSK signal set admits a binary isometric labeling, and the corresponding four-way partition $U(S)/U(S') = (V \cdot R_4)/R_2$

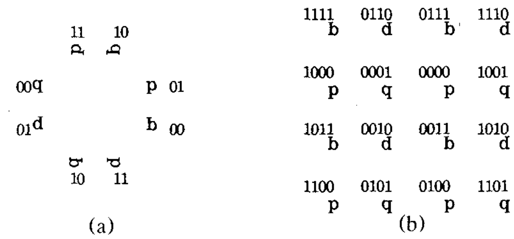

Fig. 9. Isometric Ungerboeck labelings for (a) four-way partition of an 8PSK constellation, and (b) 16-way partition of the half-integer grid.

can be refined into a chain of two two-way partitions $(V \cdot R_4)/R_4/R_2$, where the corresponding minimum squared distances are $d^2_{\min}(S_0) = 0.5858$, $d^2_{\min}(S_1) = 2$, and $d^2_{\min}(S_2) = 4$ (if the constellation points have norm 1). A corresponding isometric Ungerboeck labeling is shown in Fig. 9(a).

Similarly, for the 16-way partition of the half-integer grid, the isometry group partition chain

$$(V^2 \cdot T(2\mathbf{Z}^2))/(R_2 \cdot T(2\mathbf{Z}^2))/ T(2\mathbf{Z}^2)/T(2R\mathbf{Z}^2)/T(4\mathbf{Z}^2)$$

is a refinement of the 16-way partition $(V^2 \cdot T(2\mathbf{Z}^2))/T(4\mathbf{Z}^2)$, which partitions the half-integer grid according to the lattice partition chain $\mathbf{Z}^2/R\mathbf{Z}^2/2\mathbf{Z}^2/2R\mathbf{Z}^2/4\mathbf{Z}^2$ (where $R\mathbf{Z}^2$ is the checkerboard lattice D_2 consisting of all points in $\mathbf{Z}^2$ of even squared norm), whose minimum squared distances are $1/2/4/8/16$. A corresponding isometric Ungerboeck labeling is shown in Fig. 9(b).

Finally, we note that if S_j and $S_{j'}$ are any two elements of a partition chain $S_0/\cdots/S_n$ that admits a binary isometric labeling, then the partition $S_j/S_{j'}$ also admits a binary isometric labeling, since the corresponding partition $A_j/A_{j'}$ in label space is a $(j'-j)$-dimensional subgroup of $A/\{0\} \simeq (\mathbf{Z}_2)^n$, so $A_j/A_{j'} \simeq (\mathbf{Z}_2)^{j'-j}$.

Corollary 2 (Existence of Isometric Labelings): A partition of a lattice-type signal set $S = \Lambda + \tau$ into cosets of a sublattice Λ' admits an isometric Ungerboeck labeling if $\mathbf{Z}^N/\Lambda/\Lambda'/4\mathbf{Z}^N$ is a lattice partition chain.

Proof: There exists a $2N$-bit isometric Ungerboeck labeling for $\mathbf{Z}^N/4\mathbf{Z}^N$; if $|\mathbf{Z}^N/\Lambda| = 2^j$ and $|\Lambda'/4\mathbf{Z}^N| = 2^{2N-j'}$, then an isometric Ungerboeck labeling for Λ/Λ' can be obtained by deleting the $2N - j'$ most significant and j least significant bits of the labeling for $\mathbf{Z}^N/4\mathbf{Z}^N$. □

Sublattices of $\mathbf{Z}^N$ that have $4\mathbf{Z}^N$ as a sublattice are called binary mod-4 lattices in [11], and partitions Λ/Λ' such that $\mathbf{Z}^N/\Lambda/\Lambda'/4\mathbf{Z}^N$ is a lattice partition chain are called partitions of depth $\mu \le 4$. Thus all partitions of depth $\mu \le 4$ admit isometric Ungerboeck labelings. Every code discussed in [11] is based on such a partition.

Of course, if $u(\alpha \boldsymbol{Z}^N)$ is any version of $\boldsymbol{Z}^N$, and $u(\alpha \boldsymbol{Z}^N)/\Lambda/\Lambda'/4u(\alpha \boldsymbol{Z}^N)$ is a lattice partition chain, then the corollary still holds.

IV. Geometrically Uniform Signal Space Codes

Given a geometrically uniform partition S/S' and an isometric labeling with label alphabet A, the specification of a signal space code $\mathbb{C}$ is completed by specifying a label code C, whose elements are sequences of labels that select a sequence of subsets of S via the isometric labeling. We shall see that if C is a group code over the label group A, then $\mathbb{C}$ is a generalized coset code and is geometrically uniform.

A. Sequence Spaces

A code C over an alphabet A is a set of sequences of elements of A. Such a code is therefore a subset of a sequence space A^I, defined as follows.

Definition 8: A *sequence space* A^I is the set of all sequences $\boldsymbol{a} = \{a_k, k \in I\}$ of elements a_k of some *alphabet* A, where the *index set* I is a subset of the integers, $I \subseteq \boldsymbol{Z}$. If I is finite, e.g., $I = \{k: 1 \le k \le N\}$, then A^I is the N-fold *Cartesian product* of A with itself, conventionally written as A^N. If I is infinite, e.g., $I = \boldsymbol{Z}$, then A^I is the infinite Cartesian product $A^{\boldsymbol{Z}}$.

If A is a group under some binary operation $*$, then the sequence space A^I is also a group under $*$, extended to sequences in the usual componentwise way; i.e., $\boldsymbol{a} * \boldsymbol{b} = \{a_k * b_k, k \in I\}$.

B. Signal Space Codes

A label code C defines a signal space code $\mathbb{C}$ as follows.

Definition 9: A *signal space code* $\mathbb{C}(S/S';C)$ is defined by a label alphabet A, an index set I, a *label code* C equal to any subset of the sequence space A^I, a *partition* S/S' of a signal set S into subsets, and a one-to-one *label map* m: $A \to S/S'$. The code $\mathbb{C}(S/S';C)$ is the disjoint union

$$\mathbb{C}(S/S';C) = \bigcup_{\boldsymbol{c} \in C} \boldsymbol{m}(\boldsymbol{c}),$$

of the subset sequences $\boldsymbol{m}(\boldsymbol{c}) = \{m(c_k), k \in I\}$, $\boldsymbol{c} \in C$; i.e., $\boldsymbol{m}(\boldsymbol{c})$ is the subset sequence selected by the label sequence $\boldsymbol{c} \in C$ via the label map m. A signal sequence $\boldsymbol{s}$ is a *code sequence* in $\mathbb{C}(S/S';C)$ if $\boldsymbol{s} \in \boldsymbol{m}(\boldsymbol{c})$ for some $\boldsymbol{c} \in C$; i.e., if $\{s_k \in m(c_k), k \in I\}$.

A signal space code $\mathbb{C}(S/S';C)$ is thus a subset of the sequence space S^I, the set of all sequences of elements of the signal set S. If S is a subset of real N-space $\boldsymbol{R}^N$, then S^I and thus $\mathbb{C}$ are subsets of the sequence space $(\boldsymbol{R}^N)^I$, the set of all sequences of real N-tuples. The Cartesian product space $(\boldsymbol{R}^N)^I$ is still a Euclidean space under the usual Euclidean distance metric, with squared distance defined by componentwise addition of component squared distances.

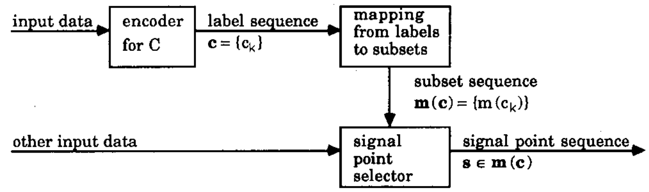

Fig. 10. Encoder for a signal space code $\mathbb{C}(S/S';C)$.

An encoder for such a code is illustrated in Fig. 10. An encoder for the label code C generates a label code sequence $\boldsymbol{c} = \{c_k\}$ that lies in C. The particular sequence $\boldsymbol{c}$ is determined by an appropriate sequence of input data. The label sequence $\boldsymbol{c}$ is then mapped to a subset sequence $\boldsymbol{m}(\boldsymbol{c}) = \{m(c_k)\}$ according to the label map m. A second operation then selects a specific signal sequence $\boldsymbol{s} \in \boldsymbol{m}(\boldsymbol{c})$ for transmission over the channel, according to a second sequence of input data. The second operation is of secondary importance; the properties of the code generated by such an encoder are primarily determined by the properties of the signal space code $\mathbb{C}(S/S';C)$, which are determined at the first level of Fig. 10.

C. Generalized Coset Codes

Now let S/S' be a geometrically uniform partition generated by a generating group partition $U(S)/U(S')$; let A be a label group isomorphic to $U(S)/U(S')$, assumed abelian with group operation $\oplus$; let the label map be the corresponding isometric labeling m: $A \to S/S'$; and let C be a group code over the label group A. That is, let C be a subgroup of the sequence space A^I, the group of all label sequences $\boldsymbol{a} = \{a_k\}$, under the extension to sequences of the group operation $\oplus$ of A. Under the assumption that A is abelian, any subgroup C of A^I is normal.

If S/S' is a partition of a lattice-type signal set $S = \Lambda + \tau$ into cosets of a sublattice Λ' of Λ induced by the translation isometry group $T(\Lambda')$, then S/S' is isomorphic to Λ/Λ', and a signal space code $\mathbb{C}(S/S';C)$ as defined here is a coset code $\mathbb{C}(\Lambda/\Lambda';C)$ in the sense of [11]. With a general geometrically uniform partition S/S', such a code may therefore be regarded as a generalized coset code.

Definition 10: A *generalized coset code* $\mathbb{C}(S/S';C)$ is the set of all signal sequences $\{\boldsymbol{s} \in \boldsymbol{m}(\boldsymbol{c}), \boldsymbol{c} \in C\}$, where S/S' is a geometrically uniform partition, C is a subgroup of the label space A^I, where A is a label group for S/S', and $\boldsymbol{m}$: $A^I \to (S/S')^I$ is the sequence extension of the corresponding isometric labeling m: $A \to S/S'$.

It is evident that $S^I/(S')^I = (S/S')^I$ is a geometrically uniform partition, and that the Cartesian product label map $\boldsymbol{m}$: $A^I \to (S/S')^I$ is an isometric labeling of this partition. For if $U(S)$ is a generating group of S, then $[U(S)]^I$ is a generating group of S^I, because any element $\boldsymbol{s} = \{s_k\}$ of S^I can be taken to any other by a composition of isometries in $U(S)$ acting on each component s_k individually. If $U(S')$ is a normal subgroup of $U(S)$, then

$[U(S')]^I$ is a normal subgroup of $[U(S)]^I$. The quotient group $[U(S')]^I/[U(S)]^I$ induces a geometrically uniform partition $S^I/(S')^I$ of S^I into subsets corresponding to cosets of $[U(S')]^I$ in $[U(S)]^I$. The label space A^I is isomorphic to $[U(S')]^I/[U(S)]^I$. The extended label map $\boldsymbol{m}$: $A^I \to (S/S')^I$ is thus isometric.

If C is a subgroup of A^I, then $\mathbb{C}(S/S';C)$ is a subgroup of S^I, and the quotient group $S^I/\mathbb{C}$ is isomorphic to A^I/C, under the group structure induced on S^I by $[U(S)]^I$. Furthermore, $(S')^I$ is a subgroup of $\mathbb{C}$ under the induced group structure, and the group partition chain $S^I/\mathbb{C}/(S')^I$ is isomorphic to $A^I/C/\{0\}^I$, in the sense that corresponding quotient groups are isomorphic: $S^I/\mathbb{C} \simeq A^I/C$, $\mathbb{C}/(S')^I \simeq C/\{0\}^I \simeq C$, and $S^I/(S')^I \simeq A^I/\{0\}^I \simeq A^I$. (This is an application of the "correspondence theorem.")

The cosets of C may be written as $C \oplus \boldsymbol{a}$, where $\boldsymbol{a}$ is any element of the coset. For each such coset $C \oplus \boldsymbol{a}$, the labeling m: $A \to S/S'$ defines a subset of S^I that we shall call a *label translate* $\mathbb{C}(S/S';C\oplus\boldsymbol{a})$ of $\mathbb{C}(S/S';C)$:

$$\mathbb{C}(S/S';C\oplus\boldsymbol{a}) = \bigcup_{c \in C} \boldsymbol{m}(c \oplus \boldsymbol{a}).$$

The label translates of $\mathbb{C}$ are the cosets of $\mathbb{C}$ in S^I under the induced group structure.

Theorem 5 (Generalized Coset Codes): If $\mathbb{C}(S/S';C)$ is a generalized coset code, then $S^I/\mathbb{C}/(S')^I$ is a chain of geometrically uniform partitions, and the label translates $\mathbb{C}(S/S';C\oplus\boldsymbol{a})$ of $\mathbb{C}$ are geometrically uniform, mutually congruent, and have a common isometry group $U(\mathbb{C})$, the subgroup of $[U(S)]^I$ that is the image of C under the label isomorphism.

Proof: Once $S^I/\mathbb{C}/(S')^I$ is established as a chain of geometrically uniform partitions, with a group structure on S^I induced by $[U(S)]^I$, Theorem 5 is a straightforward corollary of Theorem 2. □

Theorem 5 is our main result. It says that any generalized coset code $\mathbb{C}$ is geometrically uniform. Moreover, it says that the set S^I of all sequences of signal points in S has an exhaustive partition into label translates of $\mathbb{C}$, each of which is congruent to $\mathbb{C}$.

Theorem 1 must then hold for a generalized coset code, or for any of its label translates into Corollary 3.

Corollary 3 (Voronoi Regions, Distance Invariance): If $\mathbb{C}$ is (a label translate of) a generalized coset code, then a) the Voronoi regions associated with any two code sequences $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{C}$ are congruent; b) the distance profile $\mathrm{DP}(\boldsymbol{s}) = \{\|\boldsymbol{s}-\boldsymbol{s}'\|^2,\ \boldsymbol{s}' \in \mathbb{C}\}$ from a fixed signal point sequence $\boldsymbol{s} \in \mathbb{C}$ to all sequences $\boldsymbol{s}' \in \mathbb{C}$ is independent of $\boldsymbol{s}$.

A generalized coset code $\mathbb{C}$ with an infinite signal set S is a regular array in sequence space. The sequence space may be finite- or infinite-dimensional. If $\mathbb{C}$ is linear, then it is a lattice in sequence space. If the subset S' spans the signal space $\boldsymbol{R}^N$, so that S' is a crystal structure in signal space, then $(S')^I$ spans $(\boldsymbol{R}^N)^I$, and $\mathbb{C}$ is a crystal structure in sequence space.

A generalized coset code $\mathbb{C}$ with a finite signal set S and a finite sequence space is a Slepian-type code in a finite-dimensional sequence space. A generalized coset code with finite S and an infinite sequence space—e.g., a geometrically uniform PSK-type trellis code—is a generalized Slepian-type code in an infinite-dimensional sequence space.

This work was originally motivated by problems in extending trellis shaping concepts [14] to nonlinear shaping codes $\mathbb{C}_s$. The corollary shows that if a shaping code $\mathbb{C}_s$ is geometrically uniform, then the Voronoi regions associated with all sequences in $\mathbb{C}_s$ have the same shape, so that the second moments of all Voronoi regions of $\mathbb{C}_s$ (or of any label translate of $\mathbb{C}_s$) are the same. Label translates of $\mathbb{C}_s$ are used in [14] to avoid error propagation by means of feedbackfree syndrome-formers.

1) Examples: Since any subgroup C of A^I generates a geometrically uniform code, it is very easy to generate examples of such codes.

For example, Fig. 8(c) is an isometric labeling of the sequence space partition $(S/S')^2 = (\boldsymbol{Z}/4\boldsymbol{Z})^2 = \boldsymbol{Z}^2/4\boldsymbol{Z}^2$ by the Cartesian product label group A^2, where $A = (\boldsymbol{Z}_2)^2$. Taking C_1 as the (4,1) binary linear code {0000, 1001}, C_2 as the (4,2) binary linear code {0000, 1001, 0100, 1101}, and C_3 as the (4,3) binary linear code generated by {0010, 1000, 0101}, we obtain the three arrays $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ illustrated in Fig. 11. Theorem 5 guarantees that not only are these arrays geometrically uniform, but also so are all of the label translates of these arrays, and the partition of $\boldsymbol{Z}^2$ into these label translates is a geometrically uniform partition.

The arrays $\mathbb{C}_1$ and $\mathbb{C}_3$ resemble the nonlattice regular arrays shown in Fig. 6, and indeed could be made equal to the arrays of Fig. 6 by starting with a Type IV one-dimensional signal set S, rather than Type III.

The array $\mathbb{C}_2$ is actually a lattice. Li [27] noticed that $\mathbb{C}_2$ has the same density as $2\boldsymbol{Z}^2$ and the same minimum distance, but $\mathbb{C}_2$ has only two nearest neighbors rather than four, so it has a slightly higher effective coding gain [11]. $\mathbb{C}_2$ also has a hexagonal Voronoi region (not a regular hexagon), with a lower second moment than $2\boldsymbol{Z}^2$, so it is superior for shaping or quantization as well. $\mathbb{C}_2$ is not quite as good as the hexagonal lattice A_2; however, unlike A_2, it is a sublattice of $\boldsymbol{Z}^2$, and indeed $\boldsymbol{Z}^2/\mathbb{C}_2/4\boldsymbol{Z}^2$ is a chain of four-way lattice partitions. This partition has been used by Calderbank and Sloane [13] in trellis code constructions.

D. Classes of Generalized Coset Codes

Many Euclidean space codes are based on linear binary codes C over the binary field $\boldsymbol{F} = \mathrm{GF}(2)$. As we have already remarked, there is no significant distinction between additive groups isomorphic to $(\boldsymbol{Z}_2)^n$ and n-dimensional linear vector spaces over $\boldsymbol{F}$, so in this section we write $\boldsymbol{F}$ for $\boldsymbol{Z}_2$.

It was noted in Section III-A that an N-fold Cartesian product of the usual binary labeling $\{0,1\} \to \{\pm 1\}$ of the

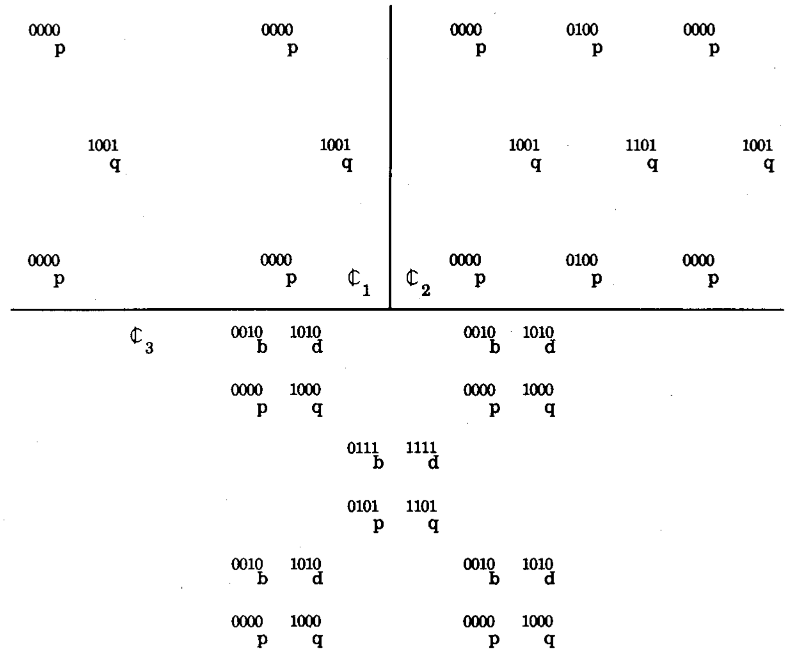

Fig. 11. Examples of generalized coset codes derived from Fig. 8(c): two regular arrays $\mathbb{C}_1$ and $\mathbb{C}_3$ as in Fig. 6, and the lattice $\mathbb{C}_2$.

binary signal set is a binary isometric labeling of the N-dimensional hypercube, using the label alphabet $\boldsymbol{F}^N$. Any additive subgroup of $\boldsymbol{F}^N$ is also a vector space of some dimension K over $\boldsymbol{F}$—i.e., an (N, K) linear binary block code C. Consequently, classical binary signal space codes $\mathbb{C}$ corresponding to any linear binary block code C are geometrically uniform. The cosets of C map into geometrically uniform signal sets that are all congruent to $\mathbb{C}$, and $\mathbb{C}$ and its cosets form a partition of the hypercube (the image of $\boldsymbol{F}^N$).

Similarly, a linear binary rate-k/n convolutional code C is a subgroup of $(\boldsymbol{F}^n)^{\boldsymbol{Z}}$, where the index set is the infinite set of all integers, $I = \boldsymbol{Z}$. Therefore the corresponding set of binary signal sequences is a geometrically uniform code $\mathbb{C}$. In practice, signal space codes for bit rates of less than one bit per dimension are often binary convolutional codes of this type. Again, the cosets of C map into geometrically uniform signal sets that are all congruent to $\mathbb{C}$, and $\mathbb{C}$ and its cosets form a partition of an infinite-dimensional hypercubic sequence space, the image of $(\boldsymbol{F}^n)^{\boldsymbol{Z}}$.

Many trellis codes $\mathbb{C}(S/S';C)$ for higher bit rates have been constructed using 2^n-way partitions S/S', n-bit binary label alphabets $A = \boldsymbol{F}^n$, and rate-k/n linear binary convolutional codes C. Such a trellis code will be a generalized coset code, and therefore geometrically uniform, if S/S' is a geometrically uniform partition that admits a binary isometric labeling, and if in fact such a labeling is used in the code.

The four-state PSK-type trellis code of Ungerboeck [12] uses the four-way partition of the 8PSK signal set shown in Fig. 9. All binary labelings of this partition are isometric, and therefore this code is geometrically uniform. There exist similar isometric labelings for any four-way partition of any 2^n-PSK signal set, and any code based on any such partition will be geometrically uniform (e.g., those of [17]). However, binary isometric labelings do not exist for partitions into more than 4 subsets, as will be shown in Section V. The remaining Ungerboeck PSK-type trellis codes use eight-way partitions of the 8PSK signal set.

The one-dimensional trellis codes of Ungerboeck [12] use the four-way partition of the one-dimensional half-integer grid shown in Fig. 8, and are therefore geometrically uniform. As with PSK, however, binary isometric labelings do not exist for 1-D partitions into more than four subsets, as will be shown in Section V.

In general, higher-dimensional lattice-type trellis codes have used chains of two-way partitions with Ungerboeck labelings. We have shown that any N-dimensional lattice partition Λ/Λ' admits a binary isometric labeling, provided that $\boldsymbol{Z}^N/\Lambda/\Lambda'/4\boldsymbol{Z}^N$ is a lattice partition chain. As already noted, all coset codes reviewed in [11] use lattice partitions of this type.

V. Other Notions of Uniformity

In this section we compare and contrast our definition of geometrical uniformity with other types of symmetry that have been previously introduced. We also use a result of Li [28] on the nonexistence of regular binary labelings for certain partitions to deduce that these partitions do not admit binary isometric labelings.

A. Linearity

For partitions Λ/Λ' of a lattice Λ into cosets of a sublattice Λ' of Λ, it is possible to have a labeling that is linear, in the following sense:

Definition 11: A labeling $m: A \to \Lambda/\Lambda'$, $m(a) = \Lambda' + t(a)$ of a lattice partition is *linear* if

$$m(a \oplus b) = \Lambda' + t(a \oplus b) = \Lambda' + t(a) + t(b)$$
$$= m(b) + t(a), \qquad a, b \in A,$$

where "+" is ordinary addition in $\boldsymbol{R}^N$.

By Proposition 2 of Section III-B, a linear labeling is an isometric labeling, with the isometry u_a being a translation by $t(a)$. This implies that $t(0)$ must be an element of Λ', so that $m(0) = \Lambda' + t(0) = \Lambda'$. Consequently there is no linear labeling for the partition of any translate of Λ into cosets of Λ', except of Λ itself.

However, if there is a linear labeling $m(a) = \Lambda' + t(a)$ for a lattice partition Λ/Λ', then the simple signal space translate $m(a) = \Lambda' + t(a) + \tau$ is an isometric labeling for the partition of the translate $\Lambda + \tau$ into cosets $\Lambda' + t(a) + \tau$ of Λ', so that partitions of lattice translates can be dealt with by simple extensions of linear techniques.

If $\mathbb{C}(\Lambda/\Lambda'; C)$ is a signal space code based on a lattice partition Λ/Λ' and a linear labeling, then $\mathbb{C}$ is itself linear, in the sense that $\mathbb{C}$ is an additive group under ordinary sequence addition in the sequence space $(\boldsymbol{R}^N)^I$. For let $\boldsymbol{s}$ and $\boldsymbol{s}'$ be any two code sequences in $\mathbb{C}$, belonging to the coset sequences $(\Lambda')^I + \boldsymbol{t}(\boldsymbol{c})$ and $(\Lambda')^I + \boldsymbol{t}(\boldsymbol{c}')$, respectively. Then the sequence $\boldsymbol{s} + \boldsymbol{s}'$ belongs to the coset sequence $(\Lambda')^I + \boldsymbol{t}(\boldsymbol{c}) + \boldsymbol{t}(\boldsymbol{c}') = (\Lambda')^I + \boldsymbol{t}(\boldsymbol{c} \oplus \boldsymbol{c}')$ and is itself in $\mathbb{C}$, since $\boldsymbol{c} \oplus \boldsymbol{c}'$ is in the linear code C.

A *lattice-type coset code* is a signal space code

$$\mathbb{C}[(\Lambda + \tau)/\Lambda'; C]$$

based on a partition of a translate $\Lambda + \tau$ of Λ into cosets of Λ'. Obviously

$$\mathbb{C}[(\Lambda + \tau)/\Lambda'; C] = \mathbb{C}(\Lambda/\Lambda'; C) + (\tau)^I,$$

so that such a code is merely a translate of a code $\mathbb{C}(\Lambda/\Lambda'; C)$ based on a partition of Λ itself. It is not strictly linear, but it can be regarded as effectively linear. Alternatively, it can be regarded as a generalized coset code with an isometric labeling, where all isometries are translations. This implies that it is geometrically uniform, which also follows from the geometric uniformity of linear codes.

A *binary linear labeling* is a linear labeling by a group isomorphic to $(\boldsymbol{Z}_2)^n$, for some integer n. The conditions under which binary linear labelings exist are rather special. The one-dimensional lattice partition $\boldsymbol{Z}/2\boldsymbol{Z}$ has a binary linear labeling by $\boldsymbol{Z}_2 = \{0, 1\}$ in which $a \in \boldsymbol{Z}_2$ maps to $2\boldsymbol{Z} + a$; i.e., $t(a) = a$, where $a \in \{0, 1\}$ is taken as an element of $\boldsymbol{Z}_2$ in the label space but as an integer in the signal space. Similarly, the Cartesian product labeling $m(\boldsymbol{a}) = 2\boldsymbol{Z}^N + \boldsymbol{a}$, $\boldsymbol{a} \in (\boldsymbol{Z}_2)^N$ of the N-dimensional lattice partition $\boldsymbol{Z}^N/2\boldsymbol{Z}^N$ by the label group $(\boldsymbol{Z}_2)^N$ is linear, and so is the corresponding labeling of any lattice partition Λ/Λ' such that $\boldsymbol{Z}^N/\Lambda/\Lambda'/2\boldsymbol{Z}^N$ is a lattice partition.

In [11], coset codes based on such partitions are called mod-2 codes, or codes of depth $\mu \le 2$, and it is shown that the two-dimensional four-state code of Ungerboeck [12] and most of the codes of Wei [15] are included in this class. Linear codes of this type are isomorphic to the binary codes C that generate them, and the minimum squared Euclidean distance between coset sequences $2\boldsymbol{Z}^N + \boldsymbol{c}$ and $2\boldsymbol{Z}^N + \boldsymbol{c}'$ is equal to the Hamming distance between $\boldsymbol{c}$ and $\boldsymbol{c}'$ as binary code sequences in C.

In general, a defining characteristic of groups that are isomorphic to $(\boldsymbol{Z}_2)^n$ is that such groups are precisely the finite groups G in which every element g is its own inverse; i.e., every $g \in G$ satisfies $g^2 = e$.

Lemma: A group G is isomorphic to $(\boldsymbol{Z}_2)^n$ for some nonnegative integer n if and only if G is a finite group such that every $g \in G$ satisfies $g^2 = e$.

Proof: If $G \simeq (\boldsymbol{Z}_2)^n$, then G is finite and abelian, and any $g \in G$ satisfies $g \oplus g = 0$. Conversely, if every element of G is its own inverse, then for any $a, b \in G$, $(ab)^2 = e$, which implies $ab = b^{-1}a^{-1} = ba$, so G is abelian. Let the identity e be denoted by 0 and the group operation be denoted by $\oplus$. Since G is abelian and every $g \in G$ satisfies $g \oplus g = 0$, any sum involving multiple occurrences of any $g \in G$ depends only on whether the number of such occurrences is odd or even. Now if $G \neq \{0\}$, choose any nonzero $g_1 \in G$; the group generated by g_1 is therefore the two-element group $G_1 = \{0, g_1\}$. If $G \neq G_1$, choose any $g_2 \in G$ not in G_1; the group generated by $\{g_1, g_2\}$ is the four-element group $G_2 = \{\Sigma_{1 \le j \le 2} a_j g_j,\ a_j \in \{0,1\},\ 1 \le j \le 2\}$. If $G \neq G_2$, continue. This process must terminate after a finite number n of iterations, since G is finite. The correspondence

$$\sum_{1 \le j \le n} a_j g_j \leftrightarrow \boldsymbol{a} \in (\boldsymbol{Z}_2)^n$$

is then an isomorphism between G and $(\boldsymbol{Z}_2)^n$. □

Therefore Λ/Λ' is isomorphic to $(\boldsymbol{Z}_2)^n$ if and only if each nonzero coset of Λ' in Λ is its own inverse under addition modulo Λ'.

Theorem 6 (Binary Linear Labelings): If Λ' is a sublattice of Λ and Λ/Λ' is finite, then there exists a binary linear labeling for Λ/Λ' if and only if 2Λ is a sublattice of Λ'.

Proof: Let $\Lambda' + t$, $t \in \Lambda$, be any coset of Λ' in Λ. Then $(\Lambda' + t) + (\Lambda' + t) = \Lambda' + 2t$ is equal to the zero coset Λ' if and only if $2t \in \Lambda'$. It follows that every coset of Λ' in Λ is its own inverse if and only if $2t \in \Lambda'$ for every $t \in \Lambda$; i.e., iff 2Λ is a sublattice of Λ'. □

If Λ is N-dimensional, then $\Lambda/2\Lambda$ has order 2^N. If $G = \{g_j, 1 \le j \le N\}$ is a generator matrix for Λ, then $2G$ is a generator matrix for 2Λ. The generators g_j are N linearly independent elements of Λ that are not in 2Λ. The correspondence

$$\left(2\Lambda + \sum_{1 \le j \le N} a_j g_j\right) \leftrightarrow \boldsymbol{a} \in (\boldsymbol{Z}_2)^N$$

is an isomorphism between $\Lambda/2\Lambda$ and $(\boldsymbol{Z}_2)^N$, because $2g_j$ is an element of 2Λ, $1 \le j \le N$. (Again, a binary variable $a_j \in \{0, 1\}$ is interpreted as an integer on the left side and as an element of $\boldsymbol{Z}_2$ on the right side of this correspondence.)

Many coset codes $\mathbb{C}(\Lambda/\Lambda';C)$ that use binary codes C are based on lattice partitions Λ/Λ' that do not admit binary linear labelings. For example, there is no binary linear labeling of the four-way partition $Z/4Z$ by $A=(Z_2)^2$, since if a is the label for the coset $4Z+1$, then $a\oplus a=00$ must map to $4Z$, not to $4Z+2$, so linearity cannot be achieved. Similarly, there exists no binary linear labeling of any lattice partition Z^N/Λ' unless $2Z^N$ is a sublattice of Λ'. In contrast, we have seen that there exists a binary isometric labeling whenever $4Z^N$ is a sublattice of Λ'.

B. The Time-Zero Subset; Ungerboeck–Zehavi–Wolf Symmetry

Let $\mathbb{C}(S/S';C)$ be a signal space trellis code, with C a rate-k/n binary linear convolutional code. Then when the encoder is in the zero state, the next encoder output can take on (at most) one of 2^k values $a\in F^n$. Since C is linear, these values form an additive group or subspace of F^n; i.e., they are the words in some (n,k) binary linear block code C_0, called the *time-zero code*. The image

$$m(C_0)=\{m(a),\, a\in C_0\}$$

of these words forms the *time-zero subset* of the signal set S, representing the set of possible signal points that could be sent from the zero state. (In [11], the corresponding sublattice $\Lambda_0=m(C_0)$ of a lattice Λ is called the "time-zero lattice.") By linearity, the set of all possible signal points that can be sent from any state (the set of "parallel transitions" corresponding to any branch) is a label translate $m(C_0\oplus b)$ of $m(C_0)$, where $C_0\oplus b$ is one of the 2^{n-k} cosets of C_0.

If $\mathbb{C}$ is a generalized coset code, then Theorem 2 applies to Theorem 7.

Theorem 7 (Time-Zero Subsets): If S/S' is a geometrically uniform partition with a binary isometric labeling m: $A\to S/S'$, where $A=(Z_2)^n$, then a) the time-zero subset $m(C_0)$ is geometrically uniform; b) the label translates $m(C_0\oplus b)$ are all geometrically congruent to $m(C_0)$; and c) the 2^{n-k}-way partition of S into the label translates $m(C_0\oplus b)$ is a geometrically uniform partition.

Theorem 7 also holds under weaker conditions. Most trellis codes use rate-$k/(k+1)$ binary convolutional codes C. In this case there are only two label translates $m(C_0\oplus b)$,the time-zero subset $m(C_0)$ and its complement [16]. If S is geometrically uniform, then it will almost always be possible to choose a labeling such that this two-way partition is geometrically uniform, whether or not an isometric labeling is possible for the full partition S/S'. Even when S is not uniform, the time-zero subset $m(C_0)$ and its complement are usually reflections of one another.

In [12], Ungerboeck assumes a rate-$k/(k+1)$ code C and a symmetrical signal set S with a label map m: $A\to S/S'$ from $A=(Z_2)^{k+1}$ in which every subset $m(a)$ is a single point in S. If the weight $w^2(b)$ is defined as

$$w^2(b)=\min_a\{d^2[m(a),m(a\oplus b)]\},\qquad ab\in(Z_2)^{k+1},$$

then Ungerboeck's symmetry condition is that this minimum be achieved for at least two values of a, one in $m(C_0)$ and one in its complement. (This concept can be generalized to arbitrary subsets by defining $w^2(b)=\min d^2_{\min}[m(a),m(a\oplus b)]$.) Then he shows that if $d^2_{\min}(\mathbb{C})$ is defined as the minimum of $w^2(c)$ over all nonzero $c\in C$, then there exist subset sequences $m(c)$ and $m(c\oplus c')$ such that the bound $d^2_{\min}[m(c),m(c\oplus c')]\ge d^2_{\min}(\mathbb{C})$ is achieved with equality.

Note that if the labeling function is actually isometric, then $m(a)=u_a[m(0)]$ and $m(a\oplus b)=u_a[m(b)]$, so $w^2(b)=\min\{d^2[m(0),m(b)]\}$, and the minimum is achieved for every a. Similarly, the bound $d^2_{\min}[m(c),m(c\oplus c')]\ge d^2_{\min}(\mathbb{C})$ is achieved for every $c\in C$, whereas under Ungerboeck's symmetry condition all that can be said is that it is achieved for some $c\in C$. (In [27], this is what distinguishes a "homogeneous" from a "strongly homogeneous" partition.)

For the same kinds of trellis codes, Zehavi and Wolf [16] define a similar but stronger notion of symmetry as follows. They define the *weight profile* of a subset such as $m(C_0)$ or its complement as the set of all squared Euclidean distances $d^2[m(a),m(a\oplus b)]$ that occur as a ranges through the subset, for a fixed b, whereas Ungerboeck [12] considers only the minimum squared distance. Their symmetry condition is that this weight profile be the same for both subsets. Then they show that the set of all distances $\{d^2[m(c),m(c\oplus c')]\}$ can be determined from the state diagram or trellis diagram for the error sequences $c\in C$. (A similar result had been shown earlier in an unpublished memorandum by Honig [29].)

Again, if the labeling function is actually isometric, then the weight profiles are all the same, and the set of distances $\{d^2[m(c),m(c\oplus c')]\}$ is equal to the set $\{d^2[m(0),m(c')]\}$ for every $c'\in C$, whereas the Ungerboeck–Zehavi–Wolf (UZW) symmetry condition allows computation of this set only over the ensemble of all $c'\in C$. On the other hand, the weaker UZW symmetry condition applies to some finite constellations that are not geometrically uniform, such as the unequally-spaced 4-PAM signal set that they use for illustration. (As in [20], this is a finite subset of a Type IV PAM set; it might be preferable to regard such a signal set as a finite subset of an infinite, geometrically uniform array, and to utilize the symmetry properties of this array.)

Biglieri and McLane [30] have pointed out that UZW symmetry follows whenever there is an isometry between the time-zero subset and its complement, a condition that is commonly met. In particular, there is such an isometry whenever this two-way partition is geometrically uniform. They also consider other metric spaces.

Buz [31] has pointed out that UZW symmetry holds only if an appropriate labeling is used, and has extended the method to general rate-k/n convolutional codes C.

He also considers certain nonlinear codes, such as a version of the eight-state two-dimensional code of Wei [32] that is used in V.32 modems.

C. Regular Labelings

Calderbank and Sloane [13] introduced trellis codes based on lattice partitions, with label maps $m(a) = \Lambda' + t(a)$, $a \in A$. Their definition of regularity is as follows.

Definition 12: A label map m: $A \to \Lambda / \Lambda'$ specified by $m(a) = \Lambda' + t(a)$ is (weak-sense) *regular* if the minimum squared distance between elements of two cosets depends only on the difference (under $\oplus$) of the coset labels:

$$d^2_{\min}[m(a), m(a')] = d^2_{\min}[\Lambda' + t(a), \Lambda' + t(a')] = N(-a \oplus a').$$

Since $d^2_{\min}[\Lambda' + t(a), \Lambda' + t(a')]$ is the squared norm (minimum squared norm of any element) of the difference coset $\Lambda' - t(a) + t(a')$, a labeling is regular if and only if these norms are all equal to each other, and in particular to the squared norm

$$N(-a \oplus a') \triangleq \mathrm{Norm}[\Lambda' - t(0) + t(-a \oplus a')].$$

A stronger notion of regularity is as follows.

Definition 12: A label map m: $A \to \Lambda / \Lambda'$ specified by $m(a) = \Lambda' + t(a)$ is *strong-sense regular* if the set of distances (distance profile) between any element of one coset $\Lambda' + t(a)$ and all elements of another coset $\Lambda' + t(a')$ depends only on the difference (under $\oplus$) of the coset labels; i.e., if the weight profile of any difference coset $\Lambda' - t(a) + t(a')$ is equal to the weight profile of the difference coset $\Lambda' - t(0) + t(-a \oplus a')$.

Obviously strong-sense regularity implies weak-sense regularity.

Property 4: An isometric labeling of a lattice partition is strong-sense regular.

Proof: If $m(a) = \Lambda' + t(a)$ is an isometric labeling, then by Proposition 2 there exists an isometry u_a such that

$$\Lambda' + t(-a \oplus a') = u_{-a}[\Lambda' + t(a')];$$
$$\Lambda' + t(0) = u_{-a}[\Lambda' + t(a)].$$

Because u_{-a} is an isometry, the set of distances from any element of $\Lambda' + t(a)$ to all elements of $\Lambda' + t(a')$ is the same as the set of distances from any element of $\Lambda' + t(0)$ to all elements of $\Lambda' + t(-a \oplus a')$, which is the same as the weight profile of the difference coset $\Lambda' - t(0) + t(-a \oplus a')$, which depends only on the label difference $-a \oplus a'$. □

As we have shown, there exist isometric and therefore strong-sense regular binary labelings for any partition Λ / Λ' such that $Z^N / \Lambda / \Lambda' / 4Z^N$ is a lattice partition.

Li [28] has shown that the only possible regular (and thus the only possible isometric) binary labelings of $Z^N / 4Z^N$ are linear transformations of the Cartesian product of an isometric labeling m: $(Z_2)^2 \to Z/4Z$, and also that there can be no regular binary labeling for any finer partition of Z^N than $Z^N/4Z^N$. The proof of the latter result is essentially as follows. Every point in Z^N has $2N$ nearest neighbors. If the labeling is regular, then the set of label differences between the label of any point and the labels of its nearest neighbors must be the same for all points. But any point in Z^N can be reached from any other point in a series of unit steps. Therefore every point in Z^N must have a label that is a linear combination over $F = Z_2$ of these $2N$ nearest-neighbor label differences, so the labels are contained in a vector space of dimension at most $2N$ over F. This means that there can be at most $2^{2N} = 4^N$ distinct labels, if the labeling is regular.

Since isometric labelings are regular, it follows that there can be no binary isometric labelings of partitions finer than $Z^N/4Z^N$. For example, there can be no binary isometric labeling of the eight-way partition $Z/8Z$.

By the same reasoning, there can be no binary isometric labeling of an eight-way partition of a 2^n-PSK signal set, or of any finer partition.

It appears that $2N$ is the largest value of n for which there exists a binary isometric labeling by $(Z_2)^n$ of an N-dimensional partition S/S', and that the only signal sets S that admit such labelings are Cartesian products of Type IV one-dimensional signal sets; however, this conjecture has not yet been proved.

D. Superlinearity

Benedetto *et al.* [17] use a different notion of linearity than that defined in Section IV-A. The *sum* of two subsets $m(a)$ and $m(b)$, which may consist only of single points, is defined as the subset $m(a \oplus b)$, where $\oplus$ is the addition operation in the label alphabet, which they assume to be mod-2 addition. Under this definition, of course, a signal space code $\mathbb{C}(S/S'; C)$ is linear whenever C is linear. However, because this addition is not ordinary addition in signal space, such a linearity property sheds no light on distance properties.

Benedetto *et al.* therefore introduce a property called *superlinearity*, which is that the distance between symbols depend only on their *sum*; i.e., only on their label sum, which is the same as the label difference when $\oplus$ is mod-2 addition. When they consider subsets that are larger than a single point, they define the distance between subsets to be the minimum distance between their elements, and then define a labeling to be "superlinear at the reduced code level" if the distance between subsets depends only on their label difference. This is the same as the Calderbank–Sloane definition of weak-sense regularity, although in [17] it is applied to a wider class of partitions S/S', and in particular to PSK-type partitions.

Benedetto *et al.* assert that the nonexhaustive partitions of 8PSK and 16-QAM signal sets that were used by Ungerboeck [12], as well as, four-way partitions of 16-PSK and 32-PSK signal sets, are superlinear at the reduced code level. The results of this paper support these assertions. The 8PSK assertion must be in error, because as we have seen there can be no regular binary labeling of an eight-way partition of the 8PSK signal set. However, the results of this paper support their other assertions.

Finally, Benedetto *et al.* recognize that distances are not the whole story. (Again, see Slepian [19] for a counterexample.) As they say [17],

> "Theorem 3 [that all distance profiles are the same] is a necessary but not sufficient condition for [the error probability to be independent of the transmitted sequence]. One should also prove that the set of directions from any code sequence (vector) to the other vectors is the same. This guarantees that the decision regions of the code sequences are congruent. It is the authors' feeling that superlinear codes do indeed have [this property]. The proof is left for further research..."

The results of this paper provide such a proof, for geometrically uniform codes. In particular, the 16-PSK and 32-PSK codes of [17] are geometrically uniform.

VI. Distance Distributions

The global distance profile $\mathrm{DP}(\boldsymbol{s})$ of a code sequence $\boldsymbol{s}$ of a signal space code $\mathbb{C}$ may be characterized by a generating function

$$g_{\mathbb{C}}(z|\boldsymbol{s}) = \sum_{\boldsymbol{s}' \in \mathbb{C}} z^{-\|\boldsymbol{s}-\boldsymbol{s}'\|^2},$$

where z is an indeterminate. In a geometrically uniform code, $\mathrm{DP}(\boldsymbol{s})$ is independent of $\boldsymbol{s}$, so one may simply write $g_{\mathbb{C}}(z)$. Often the sequence $\boldsymbol{s}' = \boldsymbol{s}$ is omitted from the sum.

If $\mathbb{C}$ is a trellis code, then $g_{\mathbb{C}}(z)$ must be defined "per unit time" by considering only sequences $\boldsymbol{s}' \neq \boldsymbol{s}$ that diverge from $\boldsymbol{s}$ for the first time at time zero; i.e., that correspond to error events starting at time zero.

Given a geometrically uniform partition S/S' with subsets $\{m(a),\ a \in A\}$, the generating function for the distances from the initial point s_0 in the subset $m(0)$ to all points in the subset $m(a)$ may be defined as

$$g_a(z) = \sum_{s' \in m(a)} z^{-\|s_0 - s'\|^2}.$$

By geometrical uniformity, this is also the generating function for the distances from any point in any subset $m(b)$ to all points in the subset $m(a \oplus b)$.

The generating function for the distances from any sequence $\boldsymbol{s}$ that lies in a subset sequence $\boldsymbol{m(b)}$ to all points in a subset sequence $\boldsymbol{m(a \oplus b)}$ is then the product

$$\begin{aligned} g_{\boldsymbol{a}}(z) &= \sum_{\boldsymbol{s}' \in \boldsymbol{m(a)}} z^{-\|\boldsymbol{s}_0 - \boldsymbol{s}'\|^2} \\ &= \prod_{k \in I} \sum_{s' \in m(a_k)} z^{-\|s_0 - s'\|^2} \\ &= \prod_{k \in I} g_{a_k}(z). \end{aligned}$$

The generating function for the distances from any sequence $\boldsymbol{s}$ to all sequences in a geometrically uniform code $\mathbb{C} = \{\boldsymbol{s} \in \boldsymbol{m(c)},\ \boldsymbol{c} \in C\}$ is then the sum

$$g_{\mathbb{C}}(z) = \sum_{\boldsymbol{c} \in C} g_{\boldsymbol{c}}(z) = \sum_{\boldsymbol{c} \in C} \left[\prod_{k \in I} g_{c_k}(z) \right].$$

Given a trellis diagram for the label code C, each branch of the trellis associated with a code symbol c_k may be labeled by the generating function $g_{c_k}(z)$. A generating function for the set of all sequences corresponding to paths that diverge from the zero state at time zero may be built up from partial generating functions $\prod_k g_{c_k}(z)$ corresponding to partial paths through the trellis. As a practical matter, only the terms corresponding to squared distances less than some upper limit $d^2_{\max}$ need be retained. This yields a procedure for calculating the terms less than $d^2_{\max}$ in the generating function $g_{\mathbb{C}}(z)$ that has complexity of the order of a Viterbi algorithm for the label code C.

Alternatively, if the label code C is specified by the state transition diagram of a finite-state machine, then closed-form expressions for the generating function $g_{\mathbb{C}}(z)$ may be obtained as in Viterbi and Omura [33], again by labeling the transitions that are associated with a code symbol c_k by the generating function $g_{c_k}(z)$.

VII. Topics for Research

There are many apparent avenues for future research, of which the following list is a sampling.

1) Extension of the Theory: The group structure of the generating groups $U(S)$ has only begun to be exploited in this paper, and it seems that the general theory of geometrically uniform codes could be developed much further. Some recent results on "signal sets matched to groups" are presented in [34].

2) Code Constructions: This paper has merely demonstrated the symmetry properties of known codes, not sought new ones. As indicated by the example under Theorem 5, given a geometrically uniform partition, it is easy to generate large numbers of geometrically uniform codes in finite- and infinite-dimensional sequence spaces. The problem is to find the good ones.

3) Exploitation of Known Results: The study of symmetries is an ancient and well-developed mathematical topic, and is important in mathematical physics. It seems likely that the literature of these fields already contains codes and theorems that may be useful in communications. For example, new Slepian-type signal sets based on known reflection (Coxeter) groups are mentioned in [34].

4) Extension of Known Results to Sequence Space: The existing mathematical and physical literature is concerned with symmetries of finite-dimensional spaces. Perhaps the usefulness in communications of geometrically uniform trellis codes in sequence spaces of countably infinite dimension will stimulate work in mathematics and physics to extend known results to infinite-dimensional spaces; e.g., to develop a theory of infinite-dimensional crystallography. Most known trellis codes are time-invariant; that is, time shifts are symmetries of the code. Time-invariant crystals in sequence space have a space-time symmetry that is an interesting generalization of the space symmetries studied in N-dimensional crystallography.

5) New Packings and Coverings: It is a bit of a mystery why the success of trellis codes has not already stimulated mathematical research into packings and coverings of infinite-dimensional sequence spaces. There may also be new nonlattice packings and coverings of finite-dimensional spaces that can be constructed as generalized coset codes; for example, we conjecture that the 16-dimensional nonlattice packing NR_{16} of [35] can be constructed in this way.

6) Label Codes over Nonbinary Groups: For certain partitions, the label group must be nonbinary. This suggests the study of linear codes (especially convolutional codes) over general groups. For instance, Massey, Mittelholzer *et al.* [36] have studied convolutional codes over $\mathbf{Z}_M$ as a ring, for use with MPSK signal sets. Some results on general linear codes over groups are reported in [37].

7) Rotational Invariance: It seems that the language of isometries is likely to be well suited to the development of a general theory of rotationally invariant codes, and to the construction of new such codes. In this case, generating groups $U(S)$ that contain the rotation group under which the code is to be invariant are likely to be useful. Some results in this area are discussed in [38].

8) Partitions of Nonrectangular Lattices: Certain good high-dimensional lattices can be constructed from partitions of lower-dimensional nonrectangular lattices, such as the hexagonal lattice A_2. Investigation of nonlattice generalized coset codes based on such partitions may be interesting. For example, the 24-dimensional Leech lattice can be constructed from a 27-way partition of A_2 [8]; we conjecture that there may be useful geometrically uniform codes based on the 27-element nonabelian group $[R_3 \cdot T(A_2)]/T(3A_2)$, which can be used in an isometric labeling of this partition.

9) Multilevel Generalized Coset Codes: Multilevel code constructions (see, e.g., [26]) both generate good codes and lend themselves to simplified multistage decoding algorithms. Under what conditions are such codes geometrically uniform?

10) Geometrically Uniform Hamming Space Codes: There may be useful constructions of nonlinear Hamming space codes as generalized coset codes, using isometries of Hamming sequence spaces. For example, we conjecture that the nonlinear but distance-invariant Nordstrom–Robinson binary code can be constructed in this way.

11) Higher-Dimensional PSK-Type Trellis Codes: If there exists no binary isometric labeling of an exhaustive partition of the 8PSK signal set, are there geometrically uniform partitions of 2^n-point signal sets on higher-dimensional spheres that admit binary isometric labelings and that can be used to construct higher-dimensional geometrically uniform PSK-type trellis codes?

12) Multiple Initial Vectors: The geometrically uniform signal sets and codes of this paper are generated as the orbits of a single initial point or sequence under a generating group. One generalization would be to consider the orbits of multiple initial points or sequences, as in [27].

13) Decoding Algorithms: The symmetry of permutation codes leads to very simple decoding algorithms [5]. Can the symmetries of more general geometrically uniform codes be exploited in decoding?

VIII. Conclusion

Ungerboeck says [12], "Good codes should exhibit regular structure." This paper has shown that most good known trellis codes have an even more regular geometrical structure than might have been expected.

Neither Slepian-type group codes nor lattice codes have come into wide use; however, trellis codes have become popular in practice. Trellis codes may therefore be regarded as the practical fruition of the attempts to exploit symmetry groups in coding that started with Slepian's "group codes for the Gaussian channel." While the inventors of trellis codes did not think in terms of isometries, they were clearly guided by symmetry principles. It is therefore not too surprising in retrospect that many of their constructions turn out to be geometrically uniform.

The results of this paper are quite basic, and probably do not begin to exploit the potential of group-theoretic methods for construction and analysis of Euclidean-space (or possibly Hamming-space) codes. The list of possibilities for future research listed in Section VII is no doubt quite incomplete.

Acknowledgment

The author is grateful to P. Algoet and M. D. Trott for basic education in group theory and in crystallography. The author also thanks E. Biglieri, G. J. Pottie, M. D. Trott, and a careful reviewer for comments on earlier versions of this paper.

References

[1] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, Apr. 1968.
[2] I. Ingemarsson, "Group codes for the Gaussian channel," in *Topics in Coding Theory (in honour of Lars H. Zetterberg)*. Berlin: Springer-Verlag, 1989, pp. 73–108.
[3] L. H. Zetterberg, "A class of codes for polyphase signals on a bandlimited Gaussian channel," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 385–395, July 1965.
[4] I. Ingemarsson, "Commutative group codes for the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 215–219, Mar. 1973.
[5] D. Slepian, "Permutation modulation," *Proc. IEEE*, vol. 53, pp. 228–236, Mar. 1965.
[6] ——, "A class of binary signaling alphabets," *Bell Syst. Tech. J.*, vol. 35, pp. 203–234, 1956.
[7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965, p. 263.
[8] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.
[9] H. Brown, R. Bülow, J. Neubüser, H. Wondratschek, and H. Zassenhaus, *Crystallographic Groups of Four-Dimensional Space*. New York: Wiley, 1978.
[10] R. L. E. Schwarzenberger, *N-Dimensional Crystallography*. London: Pitman, 1980.
[11] G. D. Forney, Jr., "Coset codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.

[12] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[13] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 177–195, Jan. 1987.

[14] G. D. Forney, Jr., "Trellis shaping," *1989 IEEE Inform. Theory Workshop*, Ithaca, NY, June 1989; submitted to *IEEE Trans. Inform. Theory*, July 1990.

[15] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 483–501, July 1987.

[16] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 196–202, Mar. 1987.

[17] S. Benedetto, M. A. Marsan, G. Albertengo, and E. Giachin, "Combined coding and modulation: theory and applications," *IEEE Trans. Inform. Theory*, vol. 34, pp. 223–236, Mar. 1988.

[18] M. A. Armstrong, *Groups and Symmetry*. New York: Springer-Verlag, 1988.

[19] D. Slepian, "On neighbor distances and symmetry in group codes," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 630–632, Sept. 1971.

[20] E. Biglieri and M. Elia, "On the existence of group codes for the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 399–402, May 1972.

[21] A. R. Calderbank and J. E. Mazo, "A new description of trellis codes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 784–791, Nov. 1984.

[22] D. Divsalar and J. H. Yuen, "Asymmetric MPSK for trellis codes," in *Proc. GLOBECOM'84*, Atlanta, GA, Nov. 1984, pp. 20.6.1–8.

[23] D. Divsalar and M. K. Simon, "Combined trellis coding with asymmetrical modulations," in *Proc. GLOBECOM'86*, New Orleans, LA, Dec. 1986, pp. 21.2.1–7.

[24] B. Grünbaum and G. C. Shephard, *Tilings and Patterns*. New York: W. H. Freeman, 1987.

[25] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. MacMillan, 1991.

[26] V. V. Ginzburg, "Multidimensional signals for a continuous channel," *Probl. Inform. Transm.* (English translation), vol. 20, pp. 20–34, Jan. 1984.

[27] E. Biglieri and M. Elia, "Multidimensional modulation and coding for band-limited digital channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 803–809, July 1988.

[28] Y. Li, "Regular labelings for trellis codes with rectangular signal constellations," in *IEEE Int. Symp. Inform. Theory*, San Diego, CA, Jan. 14–19, 1990; see also "Performance study and improvement of Ungerboeck-type trellis codes," M.S. thesis, Mass. Inst. Tech., Cambridge, MA, Dec. 1988.

[29] M. L. Honig, "Optimization of multilevel trellis codes using an error probability criterion," Bellcore Tech. Mem. TM-ARH-000-951, Mar. 1985.

[30] E. Biglieri and P. J. McLane, "Uniform distance and error probability properties of TCM schemes," *ICC'89 Convention Record*, pp. 1367–1372, 1989.

[31] R. Buz, "Signal mapping and nonlinear encoder for uniform trellis codes," submitted for publication.

[32] L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space—II: Nonlinear codes," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 672–686, Sept. 1984.

[33] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[34] H.-A. Loeliger, "Signal sets matched to groups," submitted to *IEEE Trans. Inform. Theory*, Nov. 1990.

[35] G. D. Forney, Jr., "Coset codes—Part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, Sept. 1988.

[36] J. L. Massey, T. Mittelholzer, T. Riedel, and M. Vollenweider, "Ring convolutional codes for phase modulation," *IEEE Int. Symp. Inform. Theory*, San Diego, CA, Jan. 14–19, 1990.

[37] G. D. Forney, Jr. and M. D. Trott, "State spaces, trellis diagrams, and minimal encoders for linear codes over groups," in *Proc. IEEE Int. Symp. Inform. Theory*, Budapest, Hungary, June 24–28, 1991, p. 45.

[38] M. D. Trott, "An algebraic framework for rotational invariance of trellis codes," in *Proc. IEEE Int. Symp. Inform. Theory*, Budapest, Hungary, June 24–28, 1991, p. 245.