Trellis Group Codes for the Gaussian Channel

Eric J. Rossin, *Member, IEEE*, Nagabhushana T. Sindhushayana, *Member, IEEE*, and Chris D. Heegard, *Fellow, IEEE*

***Abstract*—In this paper, *trellis group codes* are introduced as an extension of Slepian group codes to codes over sequence spaces. A trellis group code is defined over $\mathbb{R}^n$ as the orbit of a bi-infinite "seed sequence", $\boldsymbol{x}_0 \in (\mathbb{R}^n)^{\mathbb{Z}}$, under an infinite, defining group of transformations. This group of transformations is generated by a symbolic system. The theory is developed by combining a nontrivial extension of the notion of an isometric labeling, with results from the theory of symbolic dynamics over groups. New results presented here include a useful characterization of uniform partitions and a symbolic dynamic classification of trellis group codes. The theory is used to develop a class of rotationally invariant, nonabelian trellis group codes for QAM modulation. It is also shown that the 8-state, rotationally invariant trellis code designed by Wei, used in the V.32 (and V.32 bis) international modem standard, belongs to this class.**

***Index Terms*—Trellis-Coded Modulation (TCM), symbolic dynamics, orbit systems, trellis group codes, rotationally invariant codes.**

I. Introduction

THIS PAPER considers the symmetries of error control codes generated by finite-state machines and used-over additive white Gaussian noise channels. The basis for this work can be found in the work of Slepian [1] and the recent work of Forney [2]. In the former work, a block code in n-dimensional Euclidean space, $\mathbb{R}^n$, is created by the selection of

1) a "seed" vector $\boldsymbol{x}_0 \in \mathbb{R}^n$, and
2) a finite group Λ of linear transforms on $\mathbb{R}^n$.

The "group code" is then the finite set of vectors defined as the orbit of the seed under the action of the group, $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$. Note that in Slepian's formulation, the group code itself is not, in general, a group.

It is natural to ask if a given collection of vectors, $\mathbb{C} \subset \mathbb{R}^n$, can be obtained via Slepian's construction, i.e., *is the code* $\mathbb{C}$ *a group code*? The answer to this question lies in the study of the symmetries of $\mathbb{C}$, and the determination of whether these symmetries are sufficiently rich as to *generate* $\mathbb{C}$. This duality of viewpoints leads one to consider two problems: the "synthesis problem," or *how does one generate "good" group codes*, and the "analysis problem," or, *how does one decide if a given code is a group code*?

Forney [2] opened these two problems to the important and much broader class of "trellis codes." For purposes of this paper, a trellis code is an infinite-dimensional code, described in terms of bi-infinite sequences, based on component block codes lying in $\mathbb{R}^n$. Forney's observations suggest many open questions, several of which are addressed in this paper.

A simple extension of a Slepian group code to the sequence domain is obtained from the direct product group and direct product code. A group $\mathcal{G}$ of linear transforms on $\mathbb{R}^n$ and a point $\boldsymbol{x}_0 \in \mathbb{R}^n$ define a Slepian group code $\mathcal{G}(\boldsymbol{x}_0)$. The bi-infinite, direct product group is described by all sequences indexed by $\mathbb{Z}$, $\mathcal{G}^{\mathbb{Z}}$, and is a group under the obvious, component-wise group operation. Similarly, the direct product code is the set of sequences $\mathcal{G}(\boldsymbol{x}_0)^{\mathbb{Z}}$. If one takes the group $\Lambda = \mathcal{G}^{\mathbb{Z}}$ and the constant seed sequence $\boldsymbol{x}_0$, then $\Lambda(\boldsymbol{x}_0)$ is a trellis group code.

One interesting feature of the direct product code is that it, and its defining group Λ, are closed under the shift operator (i.e., if one takes any sequence from the code and shifts it left or right, then the result is a member of the code). In general, a nontrivial trellis group code is a shift-invariant subcode of the direct product code; it is defined from a shift-invariant subgroup $\Lambda < \mathcal{G}^{\mathbb{Z}}$ of the direct product group operating on the constant seed sequence $\boldsymbol{x}_0$. This combination produces the trellis group code $\Lambda(\boldsymbol{x}_0)$.

An elementary trellis group code, based on an n-dimensional Slepian group code, is created by the selection of

1) a "seed" vector $\boldsymbol{x}_0 \in \mathbb{R}^n$, and
2) a subgroup of the bi-infinite direct product group $\Lambda < \mathcal{G}^{\mathbb{Z}}$.

The finite group $\mathcal{G}$ of linear transforms on $\mathbb{R}^n$ generates a constituent block group code. This elementary type of trellis group code, $\Lambda(\boldsymbol{x}_0)$, models, e.g., the types of trellis coding incorporated in most power-limited applications such as satellite systems where QPSK modulation and binary convolutional coding are employed.

However, many band-limited systems, such as telephone channels and digital TV channels, incorporate trellis codes based on higher density signal sets, most notably QAM modulation. Through an observation attributed to Calderbank [2], [3], these signal sets can be viewed as the intersection of a bounding region with an infinite block group code. In this case, the constituent block group code is generated by an infinite group of translations and linear transformations, i.e.,

Manuscript received March 28, 1994; revised March 9, 1995. This work was supported in part by the NSF under Grants NCR-8903931 and NCR-9207331 and by the United States Army Research Office through the Army Center of Excellence for Symbolic Methods in Algorithmic Mathematics (ACSyAM), Mathematical Sciences Institute of Cornell University, under Contract DAAL03-91-C-0027. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, San Antonio, TX, Jan. 1993.

E. J. Rossin was with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with Next Level Communications, Rohnert Park, CA USA.

N. T. Sindhushaayana was with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853. He is now with QualComm, Inc., San Diego, CA 92121 USA.

C. D. Heegard is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA.

IEEE Log Number 9413883.

groups formed from isometries of $\mathbb{R}^n$. This observation brings the class of codes based on lattice signal sets [4]–[6] into the realm of trellis group codes.

To properly describe trellis group codes in the QAM environment, one must feature Ungerboeck's "method of set partitioning" [7], [8] and similar partitioning ideas. In this case, Forney introduced the notion of a "geometrically uniform partition." In this paper, this notion is extended in a natural way from Slepian group codes.

From the synthesis point of view, a uniform partition arises from

1) a "seed" group code $\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0)$, and
2) a group Λ of linear transforms on $\mathbb{R}^n$ with $\Gamma < \Lambda$ as a subgroup.

(There are certain quantifiable restrictions on the pair of groups Γ and Λ.) The uniform partition is then the finite set of disjoint block group codes defined from the orbit of the seed group code $\mathbb{C}_0$ under the action of the group Λ. The restrictions on the groups Γ and Λ are such that the universal group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ is partitioned into a finite number of disjoint group codes, each code being congruent to the seed group code $\mathbb{C}_0$. From the analysis viewpoint, a partitioning of a block group code is uniform if the cells of the partition are congruent group codes and the symmetries of the partition, i.e., the symmetries of the universal group code $\mathbb{C}$ that respect the partition, are rich enough to generate the partition. Note that a group code can be considered a special case of a uniform partition under the restriction that the seed group code is the singlet on set $\mathbb{C}_0 = \{\boldsymbol{x}_0\}$.

The complete theory of uniform partitions is derived in this paper. The development depends on the classical theory of "group actions on blocks" [9]. An open problem that is solved as a consequence in this treatment is the complete classification of the set of geometrically uniform partitions given nested group codes $\mathbb{C}_0 \subset \mathbb{C}$. Once a uniform partition has been established, a finite group, called the "partition permutation group," is defined by the set of permutations on the partition obtained from the symmetries of the partition. This group is often conveniently described by a finite labeling of the cells of the partition. The partition permutation group is then described in terms of the group of permutations on the labels, $\mathcal{G}_l$, called the "label group"; often the label group has a natural algebraic structure. Note that this is a logical extension of Forney's definition where the cells, or the labels of the cells, themselves form a group; under our definition the cardinality of the label group can be larger than the number of cells. This turns out to be an important attribute of our formulation related to the similar fact that the symmetry group of a group code is generally larger than the code itself.

A general trellis group code, based on an n-dimensional uniform partition of a group code, is created by the selection of

1) a "seed" group code $\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0) \subset \mathbb{R}^n$ that forms a cell of a uniform partition, and
2) a shift-invariant subgroup of the bi-infinite direct product group $\Lambda < \mathcal{G}_l^{\mathbb{Z}}$ where $\mathcal{G}_l$ is the finite label group of the uniform partition.

The code is then defined as the set of sequences described by $\Lambda(\mathbb{C}_0)$ (i.e., the set of all bi-infinite sequences drawn from the direct product group code $\mathbb{C}_0^{\mathbb{Z}}$ and permuted to congruent cells according to group sequences in Λ). Of course, in practice, if the group code $\mathbb{C}_0$ is infinite, i.e., the group Γ incorporates translations, then a bounding region is intersected with the elements of the uniform partition to ensure that each cell has an equal, finite number of elements.

A vital component in the study of trellis group codes is to understand the nature and structure of the shift-invariant subgroups of a bi-infinite direct product group $\Lambda < \mathcal{G}^{\mathbb{Z}}$. There seems to be two basic, interrelated, approaches, one based on symbolic dynamics (ergodic theory over finite sets) [10]–[14] and a second derived from a modern view of linear system theory [15]–[18]. This paper considers shift-invariant subgroups using the former approach. While most work in coding via symbolic dynamics does not involve a group structure, Kitchens studied the fundamental ideas in [13]. In the present paper, these ideas are explained and extended for application to the trellis group code problem. In particular, a classification of "symbolic dynamic groups" or "group systems," i.e., shift-invariant subgroups of the bi-infinite direct product group, and "orbit systems," i.e., the orbit of a seed under a group system, are presented.

Of special interest is the fact that group systems have a host of special properties, a subset of which are inherited by orbit systems. Furthermore, an example demonstrates the interesting fact that the minimal presentation of a group system, i.e., the smallest number of states required to describe the group, may be strictly larger than the minimal presentation of an orbit system derived from the group. This fact means, for example, that in the analysis problem for a given trellis code, one must consider symmetries of the code generated by groups of larger (state) complexity than the code itself.

Once the critical components of the theory of trellis group codes, including block group codes, uniform partitions, group systems, and orbit systems are developed, the paper presents applications of the theory to the problem of rotationally invariant trellis coding. First, the elementary trellis group codes with PSK modulation are considered. Then, the problem of rotationally invariant coding for QAM modulation, using general trellis group codes, is studied. In the process of this later development, a class of rotationally invariant trellis group codes are described that include the popular V.32/V.32 bis code as a special case. This result, which is consistent with the independent analysis of this code by Trott [16], shows that this "nonlinear" code in fact is a trellis group code and therefore inherits the attributes of this fascinating class.

The paper is organized as follows: Section II gives an overview of Slepian's group codes for the Gaussian channel, and introduces the idea of a generating set of isometries for a code. Section III discusses geometrically uniform partitions of geometrically uniform signal sets and defines isometric labelings of these partitions. Section IV introduces some basic ideas from symbolic dynamics, which is the study of shift-invariant sequences over a finite alphabet, and describes the particular structure one gets when the finite alphabet is in fact a group. Orbit systems are classified in the realm of

general symbolic systems, and relations are drawn between techniques from dynamical system theory applied to codes over groups [17], [18], and those from symbolic dynamics. Section V shows how the concepts above can be combined to describe trellis group codes in a natural manner, and presents rotationally invariant trellis codes as an application of this theory. A knowledge of basic group theory is assumed, as can be found in any introductory text, e.g., [19]–[21].

II. Block Group Codes

A. Definition of Group Codes

In 1968, Slepian [1], [22] introduced the original concept of a *group code for the Gaussian channel*. His idea was to consider the orbit of a point $\boldsymbol{x} \in \mathbb{R}^n$ under a finite group Λ of orthogonal transformations of $\mathbb{R}^n$, $\mathbb{C} = \Lambda(\boldsymbol{x})$. In this paper we broaden the notion of group codes so as to provide a common framework for the study of both finite and infinite block group codes as well as geometrically uniform trellis codes and, more generally, trellis group codes. A Slepian code is a finite isometry group code.

We begin with a few definitions and the notation that we will be using through out this paper. Let G be any group of invertible maps (transformations) from a nonempty set S to itself. We say that a subgroup of transformations $H < G$ is *transitive* on a subset $A \subseteq S$ if for any two points $a, b \in A$, there is a transformation $h \in H$ such that $h(a) = b$; H is said to be *sharply transitive* if such h is unique for each a, b (in which case $|H| = |A|$).

We define the *group of symmetries* of a subset $A \subseteq S$ (with respect to G) as the subgroup of all transformations in G under which the set A is invariant

$$\mathcal{S}\mathrm{ym}\,(A) \equiv \mathcal{S}\mathrm{ym}_G(A) \equiv \{g \in G \mid g(a) \in A, \forall a \in A\}.$$

We also define the *stabilizer* of a point $a \in S$ to be the subgroup

$$\mathcal{S}\mathrm{tab}\,(a) \equiv \mathcal{S}\mathrm{tab}_G(a) \equiv \{g \in G \mid g(a) = a\}$$

and the stabilizer of a subset A to be the intersection

$$\mathcal{S}\mathrm{tab}\,(A) = \mathcal{S}\mathrm{tab}_G(A) = \bigcap_{a \in A} \mathcal{S}\mathrm{tab}\,(a).$$

We note that the stabilizer of a point a coincides with the group of symmetries of the singleton $\mathcal{S}\mathrm{tab}\,(a) = \mathcal{S}\mathrm{ym}\,(\{a\})$, and that the stabilizer $\mathcal{S}\mathrm{tab}\,(A)$ of a set A is a *normal* subgroup of the symmetry group $\mathcal{S}\mathrm{tab}\,(A) \lhd \mathcal{S}\mathrm{ym}\,(A)$. (Recall that a subgroup $H < G$ is *normal*, written $H \lhd G$, if $g^{-1}Hg = H$ for all $g \in G$.)

A *code* over a set $\mathbb{X}$ is a (nonempty) subset $\mathbb{C} \subseteq \mathbb{X}^{\mathbb{T}}$, indexed by the set $\mathbb{T}$ (often associated with a time axis)

$$\mathbb{X}^{\mathbb{T}} \equiv \{\boldsymbol{x} \mid \boldsymbol{x} : \mathbb{T} \to \mathbb{X}, x_k \in \mathbb{X}, \forall k \in \mathbb{T}\}.$$

An element $\boldsymbol{x} \in \mathbb{C}$ is called a *codeword*. The code $\mathbb{C}$ is said to be *finite* or *infinite* depending on the cardinality $|\mathbb{C}|$ of the set $\mathbb{C}$. Suppose Σ is a group of invertible transformations of the set $\mathbb{X}^{\mathbb{T}}$. If there exists a subgroup $\Lambda < \Sigma$ and a point $\boldsymbol{x}_0 \in \mathbb{X}^{\mathbb{T}}$ such that $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$, then $\mathbb{C}$ is said to be a *group code*[1] with a *defining group* Λ and a *seed* $\boldsymbol{x}_0$. Note that the defining group and the seed of a given group code need not be unique. If the code has a defining group Λ that is finite, then $\mathbb{C}$ is a *finite group code*. If the set $\mathbb{X}^{\mathbb{T}}$ is a metric space (such as $\mathbb{R}^n$) and the ambient group of transformations Σ consists of isometries of $\mathbb{X}^{\mathbb{T}}$ (i.e., distance preserving transformations $\|\boldsymbol{x} - \boldsymbol{y}\| = \|\lambda(\boldsymbol{x}) - \lambda(\boldsymbol{y})\|$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}^{\mathbb{T}}$), then the group code $\Lambda(\boldsymbol{x}_0)$ is called an *isometry group code*.

A *generator set* for a group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ with respect to the seed $\boldsymbol{x}_0 \in \mathbb{C}$ is a minimal set of transformations $\Lambda_g \subseteq \mathcal{S}\mathrm{ym}\,(\mathbb{C})$ such that $\Lambda_g(\boldsymbol{x}_0) = \Lambda(\boldsymbol{x}_0)$.[2] If Λ is any defining group of a group code $\mathbb{C}$ with respect to (w.r.t.) a seed $\boldsymbol{x}_0$, then a generator set for $\mathbb{C}$ may be obtained by choosing a representative for each left coset[3] of the subgroup $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) < \Lambda$. Different choices of coset representatives and defining groups lead to different generating sets of a given group code.

If a generator set Λ_g of a group code $\mathbb{C}$ (w.r.t a seed x_0) is itself closed under the group operation, then it is said to be a *generating group*. In other words, a defining group of a group code that is sharply transitive ($|\Lambda| = |\mathbb{C}|$) is called a *generating group*. Note that a group code may or may not have a generating group.

Given a group code $\Lambda(\boldsymbol{x}_0)$, the stabilizer $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) < \Lambda$ is typically a nontrivial subgroup that is not normal in the defining group Λ. Nevertheless, it is often the case that one *can* chose a system of coset representatives for the left cosets that also form a group ($\Lambda_g < \Lambda$). Moreover, the resulting generating group is a normal subgroup of the defining group, $\Lambda_g \lhd \Lambda$. In this case, the group Λ has a *semidirect product decomposition* where one can write the defining group as the semidirect product of the generating group and the stabilizing group [20], $\Lambda = \Lambda_g \rtimes \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$ (see the Appendix for a discussion of semidirect products).

B. Slepian Group Codes

A *block code* $\mathbb{C}$ over $\mathbb{R}$ with *blocklength* n consists of a set of vectors in $\mathbb{R}^n$ that *span* $\mathbb{R}^n$. For a block code we can take the index set $\mathbb{T} = \{1, 2, \cdots, n\}$ and $\mathbb{C} \subset \mathbb{R}^{\mathbb{T}}$. (Note that if a block code $\mathbb{C} \subseteq \mathbb{R}^n$ does not span $\mathbb{R}^n$, then with a suitable choice of basis, $\mathbb{C}$ may be considered as a block code over $\mathbb{R}$ with block length m, where m is the dimension of the span.)

We may construct a block group code over $\mathbb{R}$ by taking Σ to be the set of all linear transformations of $\mathbb{R}^n$. A finite block group code $\Lambda(\boldsymbol{x}_0)$ is then specified by a finite group $\Lambda < \Sigma$ and a point $\boldsymbol{x}_0 \in \mathbb{R}^n$. (With somewhat of an abuse of notation, the group of linear transformations Λ is represented in two forms: in "function notation" by $\{\lambda(\cdot) \mid \lambda\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n\}$ and in "matrix notation" by $\{M \mid M \in \mathbb{R}^{n \times n}\}$, where $\lambda(\boldsymbol{x}) = M_\lambda \boldsymbol{x}$.)

[1] We adopt a terminology similar to Slepian's original use [1], where the codewords are "over the reals $\mathbb{R}$" and "group" is an adjective for the code, as in, a Reed–Solomon code is a "linear" code "over $\mathcal{F}_q$." In this language, the "group codes" in [18] are "group codes over groups" (the codewords are composed from group elements).

[2] Note that the "generator set" of a group code is a concept strictly different from the standard notion of a "generating set" for a group.

[3] A left coset of a subgroup of transformations $\mathbb{H} < \Lambda$ is represented by $\lambda_i\mathbb{H} \equiv \{\lambda_i(\lambda(\cdot)) \mid \lambda(\cdot) \in \mathbb{H}\}$, $\lambda_i \in \Lambda$ (i.e., λ_i is applied *after*). Right cosets, $\mathbb{H}\lambda_i$ are similarly defined (i.e., λ_i is applied *before*).

Since the defining group Λ is finite, every transformation $\lambda \in \Lambda$ must be of *finite order* (i.e., $\lambda^m = \boldsymbol{I}$, for some integer m, where $\boldsymbol{I}$ is the identity transformation). Hence we may confine ourselves to the situation where Σ consists only of linear operators whose determinants are ± 1.

Slepian Group Codes [1], [22] are finite block isometry group codes over the reals $\mathbb{R}$, constructed by taking the image of a point $\boldsymbol{x}_0 \in \mathbb{R}^n$ under a finite group Λ of *orthogonal* transformations of $\mathbb{R}^n$ (i.e., $M_\lambda^t M_\lambda = \underline{I}$). Since orthogonal transformations preserve the Euclidean norm of vectors, a Slepian group code is defined as points on the surface of an n-sphere of radius $\|\boldsymbol{x}_0\|$ in $\mathbb{R}^n$. Note that the code depends critically on both the choice of the defining group of transformations Λ and the seed vector, $\boldsymbol{x}_0$ [22]. When the block length n is small, the Euclidean space code is often considered a *signal constellation* or a *signal set* (see Example 1).

C. Isometry Group Codes

Every isometry of $\mathbb{R}^n$ is an affine transformation $\lambda(\boldsymbol{x}) = M_\lambda \boldsymbol{x} + \boldsymbol{c}_\lambda$, where M_λ is an orthogonal matrix [23, p. 372]. An isometry is said to be a *pure translation* if $\lambda(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{c}_\lambda$ ($M_\lambda = \boldsymbol{I}$). It is *linear* if $\lambda(\boldsymbol{x}) = M_\lambda \boldsymbol{x}$ ($\boldsymbol{c}_\lambda = \boldsymbol{0}$). The set of all isometries of $\mathbb{R}^n$ forms a group under composition denoted $\mathcal{I}\mathrm{so}_n$. The *translation group*,

$$\mathcal{T}_r \equiv \{\lambda \in \mathcal{I}\mathrm{so}_n \mid \lambda(\cdot) - \lambda(\boldsymbol{0}) = \boldsymbol{I}_n\}$$

is a normal subgroup, $\mathcal{T}_r \lhd \mathcal{I}\mathrm{so}_n$, and its quotient group, $\mathcal{I}\mathrm{so}_n/\mathcal{T}_r$, called the *linear constituent group* [2], is isomorphic to the group of all orthogonal transformations of $\mathbb{R}^n$. Given any isometry $\lambda(\boldsymbol{x})$, the *translation component* is the map $\lambda_T(\boldsymbol{x}) = \boldsymbol{x} + \lambda(\boldsymbol{0})$ and the *linear component* is the map

$$\lambda_L(\boldsymbol{x}) = \lambda(\boldsymbol{x}) - \lambda(\boldsymbol{0}) \quad (\lambda(x) = \lambda_T(\lambda_L(\boldsymbol{x}))).$$

Slepian only considered finite defining groups $\Lambda < \mathcal{I}\mathrm{so}_n$ in order to ensure a finite isometry group code $\Lambda(\boldsymbol{x}_0)$. Thus the symmetry group of a Slepian group code must have a trivial translation subgroup Λ_T. However, an alternate method of designing a finite block code over $\mathbb{R}$ begins by considering nontrivial (and therefore infinite) translation subgroups combined with a finite linear constituent group. Such an infinite defining group, Λ, generates an *infinite block isometry group code* $\Lambda(\boldsymbol{x}_0)$. Then the finite code,

$$\Lambda(\boldsymbol{x}_0)|_{\mathcal{R}} = \mathcal{R} \cap \Lambda(\boldsymbol{x}_0)$$

is a *subcode* of $\Lambda(\boldsymbol{x}_0)$ obtained by intersection with a finite volume *bounding region*, $\mathcal{R}$ (this view is attributed by Forney [2], as well as Biglieri and Elia [3] to Calderbank). Note that while $\Lambda(\boldsymbol{x}_0)|_{\mathcal{R}}$ is not a group code, it often inherits many desirable properties of the infinite group block code $\Lambda(\boldsymbol{x}_0)$, in which it lies.

D. Properties of Group Codes

Group codes have several important and distinctive properties [1], [2], [22]. The following properties apply to both finite or infinite block group codes, as well as to trellis group codes (to be defined later):

Fact 1: The code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ is invariant under Λ.
By definition, $\Lambda < \mathcal{S}\mathrm{ym}_\Sigma(\mathbb{C})$. □

Fact 2: Any point in $\Lambda(\boldsymbol{x}_0)$ can be used as the seed.
If $\boldsymbol{x}_1 \in \Lambda(\boldsymbol{x}_0)$, then $\Lambda \boldsymbol{x}_1 = \Lambda(\boldsymbol{x}_0)$. Thus a group code $\mathbb{C}$ is determined by its generating group Λ and any element $x \in \mathbb{C}$. This implies that in a group code, all the codewords are on an equal footing. □

Fact 3: A code $\mathbb{C}$ over $\mathbb{X}$ is a group code with respect to Σ if and only if (iff) its symmetry group $\mathcal{S}\mathrm{ym}_\Sigma(\mathbb{C})$ is transitive on $\mathbb{C}$.

If Λ is any subgroup of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ that is transitive on the code $\mathbb{C}$, then $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ for any $\boldsymbol{x}_0 \in \mathbb{C}$. Conversely, if $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$, then Λ is a subgroup of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ and is transitive on $\mathbb{C}$. □

Fact 4: The cardinality of a group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ divides the cardinality of the defining group Λ.

Two transformations $\lambda_1, \lambda_2 \in \Lambda$ map the seed $\boldsymbol{x}_0$ to the same point $x \in \mathbb{C}$ iff they belong to the same left coset of the stabilizer group $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$. This establishes a one-to-one correspondence between the codewords and the left cosets of $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$, denoted $= [\Lambda; \lambda_i \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)]$:[4] hence the conclusion follows from Lagrange's Theorem. □

A generator set $\Lambda_g \subseteq \Lambda$ of a group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ with respect to the seed $\boldsymbol{x}_0$ is obtained by selecting a representative from each left coset of $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$. Typically, the stabilizer subgroup $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$ is *not* normal in Λ, unless the stabilizer subgroup is trivial. To see why this is so for block codes over $\mathbb{R}^n$, consider the following argument. If $\mathcal{S}\mathrm{tab}(\boldsymbol{x}_0)$ is a normal subgroup of Λ then it is the stabilizer subgroup of every element of $\Lambda(\boldsymbol{x}_0)$. (For any point $\boldsymbol{x}_1 \in \Lambda(\boldsymbol{x}_0)$, $\boldsymbol{x}_1 = \lambda_1(\boldsymbol{x}_0)$, the stabilizer is a conjugate subgroup $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_1) = \lambda_1 \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)\lambda_1^{-1} = \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$.) Since it is assumed that the code spans $\mathbb{R}^n$, one can find a basis for $\mathbb{R}^n$ in $\Lambda(\boldsymbol{x}_0)$, and so $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$ is the stabilizer of all of $\mathbb{R}^n$ and thus trivial (i.e., $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) = \{\boldsymbol{I}\}$).

As remarked earlier, it is usually the case that there exists a generating group Λ_g, which is often normal in the defining group Λ ($\Lambda_g \lhd \Lambda$). We then have a semidirect product decomposition, $\Lambda = \Lambda_g \rtimes \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0)$.

Fact 5: Isometry Group Codes over $\mathbb{R}^n$ are Geometrically Uniform.

A code $\mathbb{C} \subseteq \mathbb{R}^n$ is *geometrically uniform* [2], [24] if for every vector, $\boldsymbol{x} \in \mathbb{C}$, the sets of difference vectors

$$\Delta(\boldsymbol{x}; \mathbb{C}) \equiv \{\boldsymbol{y} - \boldsymbol{x} \mid \boldsymbol{y} \in \mathbb{C}\}$$

are related by a rotation or reflection (more precisely, given any pair $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}$ in the code, there is an orthogonal transformation of $\mathbb{R}^n$ that maps $\Delta(\boldsymbol{x}_1; \mathbb{C})$ onto $\Delta(\boldsymbol{x}_2; \mathbb{C})$). (In a geometrically uniform universe, the stellar constellation viewed from each star is the same, or a mirror image, independent of the choice of stellar system.)

For an isometry group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0) \subseteq \mathbb{R}^n$, take $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}$ and an isometry $\lambda \in \Lambda, \lambda(\boldsymbol{x}_1) = \boldsymbol{x}_2$. Then the

[4] We denote a *partition of a set* $[A; A_i]$, where $A = \cup_i A_i$, $A_i \cap A_j = \phi$, $i \neq j$. For example, the left cosets of a subgroup $\mathbb{H} < \Lambda$ forms the partition $\Lambda\lambda_i\mathbb{H}$, the right cosets $[\Lambda; \mathbb{H}\lambda_i]$. We refer to the disjoint subsets $\{A_i\}$ as the *cells* of the partition.

linear component λ_L relates the sets

$$\Delta(\boldsymbol{x}_2;\mathbb{C}) = \lambda_L(\Delta(\boldsymbol{x}_1;\mathbb{C}))$$

and so it is easily seen to be geometrically uniform.

Conversely, given a geometrically uniform code $\mathbb{C} \subset \mathbb{R}^n$, for any two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}$, there is an orthogonal transformation γ such that

$$\Delta(\boldsymbol{x}_2;\mathbb{C}) = \gamma(\Delta(\boldsymbol{x}_1;\mathbb{C})).$$

Then the map

$$\lambda(\boldsymbol{x}) = \gamma(x - \boldsymbol{x}_1) + \boldsymbol{x}_2$$

defines an isometry of $\mathbb{R}^n$ that moves the point $\boldsymbol{x}_1$ to $\boldsymbol{x}_2 = \lambda(\boldsymbol{x}_1)$ and leaves the code invariant, $\mathbb{C} = \lambda(\mathbb{C})$. Thus $\mathbb{C}$ is an isometry group code. □

This fact may be used to generalize the notion of geometrical uniformity to spaces other than $\mathbb{R}^n$; a code $\mathbb{C}$ that lies in a metric space is geometrically uniform if its symmetry group $\mathcal{S}\mathrm{ym}_{\mathcal{I}\mathrm{SO}}(\mathbb{C})$ is transitive [24].

Geometrical uniformity implies that the *maximum-likelihood decision regions*, under additive white Gaussian noise (i.e., nearest neighbor decision regions under Euclidean distance; also called the *Voronoi regions*)

$$\mathcal{R}_V(\boldsymbol{x}) \equiv \{\boldsymbol{y} \in \mathbb{R}^n \| \boldsymbol{y} - \boldsymbol{x}\| \le \|\boldsymbol{y} - \boldsymbol{z}\|, \forall \boldsymbol{z} \in \mathbb{C}\}, \boldsymbol{x} \in \mathbb{C}$$

are also *geometrically congruent* (related by an isometry). Slepian called this symmetry condition the "equipunctional" property of group codes. This property produces a uniform decoder error probability on additive white Gaussian noise channels with maximum-likelihood decoding.

Geometric uniformity also implies that the set of distances (using the Euclidean metric) from one codeword to all others (i.e., the distance profile) is independent of the "reference" codeword. In particular, this implies that the *minimum distance* of the code

$$\begin{aligned} d_{\min} &\equiv \min_{\boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{C}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \\ &= \min_{x \in \mathbb{C}, 0 \neq z \in \Delta(x;\mathbb{C})} \|z\| \\ &= \min_{0 \neq z \in \Delta(\boldsymbol{x}_0;\mathbb{C})} \|z\| \end{aligned}$$

where $\|\cdot\|$ is the Euclidean distance. Many other symmetry properties hold, as described by Forney [2].

E. Examples of Block Group Codes

We first examine certain signal sets generated by the *dihedral group*, $\mathbb{D}_m$, the symmetries of a regular m-gon. As a group of $|\mathbb{D}_m| = 2m$ elements, $\mathbb{D}_m$ can be represented as a matrix group generated by the two linear transformations R_m and S $(= S_{\pi/4})$

$$\mathbb{D}_m \cong \langle R_m, S\rangle = \{R_m^j S^i \mid 0 \le j < m, i \in \{0,1\}\}$$

where, for $\theta = 2\pi/m$

$$R_m \equiv \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

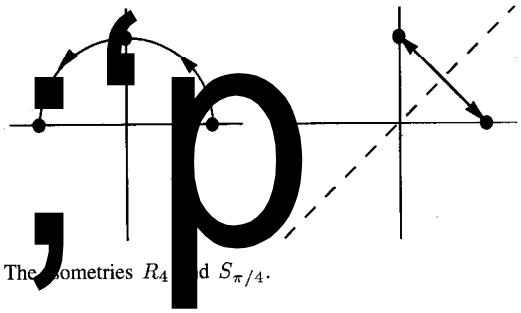

Fig. 1. The isometries R_4 and $S_{\pi/4}$.

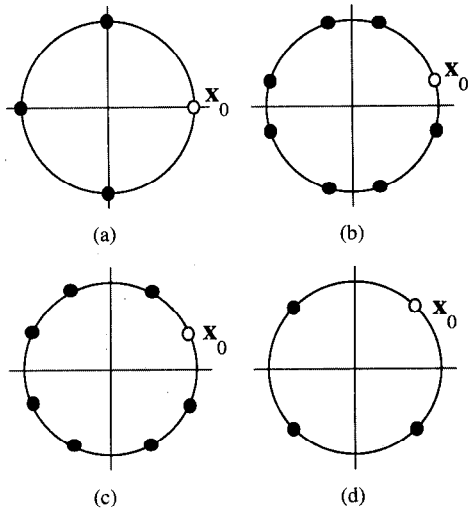

Fig. 2. Slepian signal sets $\mathbb{D}_4(\boldsymbol{x}_0)$, $\boldsymbol{x}_0 = [\cos(\phi) \quad \sin(\phi)]^t$. (a) $\phi = 0$(QPSK). (b) $\phi = \pi/16$. (c) $\phi = \pi/8$ (8-PSK). (d) $\phi = \pi/4$ (QPSK).

and

$$S_p \equiv \begin{pmatrix} \cos(2p) & \sin(2p) \\ \sin(2p) & -\cos(2p) \end{pmatrix}.$$

The isometry R_m is a pure rotation of the plane by angle θ, while S_p is a reflection about a line at an angle p.

The dihedral group satisfies the relations, $R_m^m = S^2 = \boldsymbol{I}$, $R_m^{m-1} S = S R_m$. The subgroups generated by R_m and S, are isomorphic to the integers modulo m, $\mathbb{Z}_m$, under addition, $\langle R_m\rangle \cong \mathbb{Z}_m$, $\langle S\rangle \cong \mathbb{Z}_2$, and $\mathbb{D}_n < \mathbb{D}_m$ iff $n \mid m$. There are many semidirect product decompositions of $\mathbb{D}_m$. For example, the subgroup $\langle R_m\rangle \lhd \mathbb{D}_m$ is normal and

$$\mathbb{D}_m = \langle R_m\rangle \rtimes \langle S\rangle \cong Z_m \rtimes Z_2.$$

Similarly

$$\mathbb{D}_{2m} = \mathbb{D}_m \rtimes \langle R_{2m} S\rangle \cong \mathbb{D}_m \rtimes Z_2.$$

Example 1 (Phase-Shift Keying): Consider the signal sets generated by $\Lambda = \mathbb{D}_4 = \langle R_4, S\rangle$, the symmetries of the square (Fig. 1). To generate a Slepian signal set from this group, choose an initial seed $\boldsymbol{x}_0 \in \mathbb{R}^2$ and apply each transformation $\lambda \in \mathbb{D}_4$ to $\boldsymbol{x}_0$. Fig. 2 shows that the size of the signal set is dependent on the choice of $\boldsymbol{x}_0$. If $\boldsymbol{x}_0$ is chosen as $[\cos(\phi), \sin(\phi)]^t$ with $\phi = 0 \bmod \pi/4$, then $|\Lambda(\boldsymbol{x}_0)| = 4$, and the signal set generated is QPSK. Otherwise, the set has eight elements, $|\Lambda(\boldsymbol{x}_0)| = 8$. To generate 8-PSK requires $\phi \neq 0 \bmod \pi/4$, $\phi = 0 \bmod \pi/8$. Otherwise, the eight points are not uniformly distributed around the circle.

For QPSK, the defining group $\Lambda = \mathbb{D}_4$, is also the symmetry group of the code $\mathcal{S}\mathrm{ym}(\mathbb{C})$, and there is a nontrivial stabilizer in Λ. For example, when $\phi = \pi/4$

$$\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) = \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_2) = \{I, S\}$$

and

$$\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_1) = \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_3) = \{I, R_4^2 S\}$$

(where $\boldsymbol{x}_i = R_4^i \boldsymbol{x}_0$). In this case, a generating group for $\boldsymbol{x}_1$ may be chosen to be either $\Lambda_g = \langle R_4 \rangle$ or $\Lambda_g = \langle R_4^3 S, R_4^2 \rangle$. Note that both $\langle R_4 \rangle$ and $\langle R_4^3 S, R_4^2 \rangle$ are normal subgroups of $\mathbb{D}_4$, and yet are not isomorphic, since $\langle R_4 \rangle \cong \mathbb{Z}_4$ while $\langle R_4^3 S, R_4^2 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. Both cases lead to a semidirect product composition, $\Lambda = \langle R_4 \rangle \rtimes \langle S \rangle$ and $\Lambda = \langle R_4^3 S, R_4^2 \rangle \rtimes \langle R_4 S \rangle$.

In the 8-PSK case, the defining group, $\Lambda = \mathbb{D}_4$, is a generating group while the symmetry group of the code is strictly larger, $\mathcal{S}\mathrm{ym}(\mathbb{C}) = \mathbb{D}_8$. Although the stabilizer in Λ is trivial, it is nontrivial in $\mathcal{S}\mathrm{ym}(\mathbb{C})$. For example, when $\phi = \pi/8$, $\mathcal{S}\mathrm{tab}(\boldsymbol{x}_0) = \langle R_8 S \rangle$. In this case, $\langle R_8 \rangle \cong \mathbb{Z}_8$ is also a generating group (not contained in the defining group) that is not isomorphic to the generating group $\Lambda = \mathbb{D}_4$.

When the angle of $\boldsymbol{x}_0$ does not generate PSK, then $\Lambda = \mathbb{D}_4$ is the only generating set; it is also the symmetry group $\mathcal{S}\mathrm{ym}(\mathbb{C})$ of the group code. □

Example 2 (Quadrature Amplitude Modulation): Consider the signal sets of Fig. 3, generated by groups that include translations. The infinite lattice signal sets of Fig. 3(a) and (b) can be generated using only translations and a seed $\boldsymbol{x}_0 = \mathbf{0}$ at the origin. For the *integer lattice*, Fig. 3(a), the points are generated by the group of translations $\Lambda = \langle T_x, T_y \rangle$, where $T_x \equiv T_{\Delta,0}$, $T_y \equiv T_{0,\Delta}$, and Δ is the minimum distance of the lattice and the general two–dimensional translation isometry

$$T_{a,b}(x, y) \equiv (x + a, y + b) \in \mathcal{T}_r.$$

Similarly, for the *hexagonal lattice*, Fig. 3(b), the points are generated by the group of translations $\Lambda = \langle T_x, T_{\Delta/2, \Delta\sqrt{3}/2} \rangle$. In both cases, the symmetries of the signal set have a nontrivial linear constituent group ($\mathbb{D}_4$ and $\mathbb{D}_6$, respectively) and can be generated by other subgroups of the symmetry group. For example, the integer lattice is generated by the groups $\Lambda = \langle R_4, T_x \rangle$ and $\Lambda = \langle S, T_x \rangle$ and the hexagonal lattice is generated by the groups $\Lambda = \langle R_6, T_x \rangle$ with the reflection about 30°, $\Lambda = \langle S_{\pi/6}, T_x \rangle$.

Fig. 3(c) shows a *regular array*, an infinite signal set which is not a lattice (i.e., the points are not closed under vector addition). The *four-point checkers array* can be generated by diagonal translations of QPSK, for example

$$\Lambda = \langle \mathbb{D}_4, T_x^2 T_y^2 = T_{2\Delta, 2\Delta} \rangle.$$

The infinite *Quadrature Amplitude Modulation* (QAM) signal set is obtained from a group Λ that generates the integer lattice with a seed $\boldsymbol{x}_0 = (\Delta/2, \Delta/2)$. Thus QAM is based on a nontrivial translate of the integer lattice. A finite QAM signal set is obtained by imposing a finite bounding region $\mathcal{R}$ (usually a square or a cross); Fig. 3(d) shows the 32 and 64 QAM signal set.

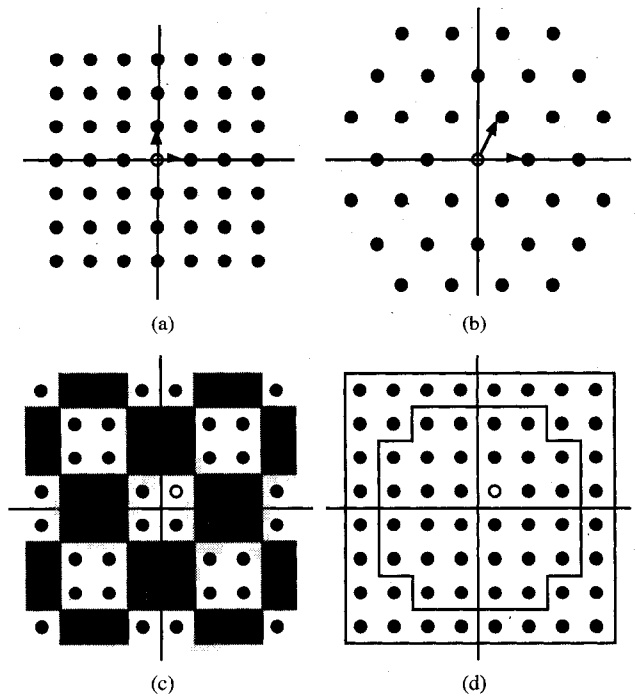

Fig. 3. Infinite signal sets. (a) Integer lattice. (b) Hexagonal lattice. (c) Four-point checkers array. (d) Integer lattice translate (QAM).

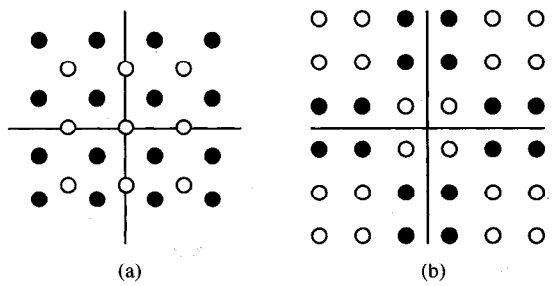

Fig. 4. Partitioning signal sets via subgroups of index 2. (a) QAM from a coset of the integer lattice. (b) QAM partitioned by the four-point checkers array.

Finally, we note that the infinite QAM signal set can also be described in terms of a natural partition of the integer lattice (Fig. 4(a)). If one generates the (rotated) integer lattice with the group

$$\Lambda = \left\langle T_{\Delta/\sqrt{(2)}, \Delta/\sqrt{(2)}}, T_{\Delta/\sqrt{(2)}, -\Delta/\sqrt{(2)}} \right\rangle$$

and the seed $\boldsymbol{x}_0 = \mathbf{0}$, then the normal subgroup $\Gamma = \langle T_x, T_y \rangle \lhd \Lambda$, of index 2, induces a partition of the lattice into two congruent signal sets. The subgroup Γ acting on the seed x_0 generates the integer sublattice while its coset generates the QAM signal set. Similarly, the four-point checkers array is associated with a two-way partition of the QAM signal set by a subgroup of index 2 (Fig. 4(b)). Such phenomena are basic examples of *geometrically uniform partitions* of a group code, which constitutes the main subject of the next section. □

III. Group Code Partitions and Labelings

The primary geometric objects (in $\mathbb{R}^n$) that will be needed for trellis coding are geometrically uniform (GU) partitions of

signal sets. In this section, we construct such partitions and develop a language that will characterize labelings that are useful for coding.

A. Groups Actions on Points, Sets, and Partitions

In the study of isometry group codes over $\mathbb{R}^n$, one deals with three basic objects: *points* $\boldsymbol{x} \in \mathbb{X}^{\mathbb{T}} = \mathbb{R}^n$, *sets* $\mathbb{C} \subset \mathbb{X}^{\mathbb{T}}$, and *partitions* $[\mathbb{C};\ \mathbb{C}_i]$; the subsets $\{\mathbb{C}_i\}$ are the *cells* of the partition. Again, for a uniform framework, consider a group G invertible transformations on a fixed set S. Let $A \subset S$ be a set, take $a \in A$ as a point of the set, let $[A; A_i]$ be a partition of the set.

The *symmetry group* of an object is the subgroup of the group G under which the object is invariant

$$\begin{aligned}\mathcal{S}\mathrm{ym}_G(a) &\equiv \{g \in G \mid g(a) = a\}\\ \mathcal{S}\mathrm{ym}_G(A) &\equiv \{g \in G \mid g(A) = A\}\\ \mathcal{S}\mathrm{ym}_G([A; A_i]) &\equiv \{g \in G \mid g([A; A_i]) \in [A; A_i]\}\end{aligned}$$

where

$$g(A) \equiv \{g(a) \mid a \in A\}$$

and

$$g([A; A_i]) \equiv \{g(A_i),\ \ A_i \in [A; A_i]\}.$$

When the group G is clear, we write $\mathcal{S}\mathrm{ym}(\cdot)$ for $\mathcal{S}\mathrm{ym}_G(\cdot)$. Note that the image of a set, $g(A)$, is a subset of S and the image of a partition, $g([A; A_i])$, is a partition of $g(A)$. Furthermore, symmetries of a set can permute the points within the set, while symmetries of a partition can permute both the points within the cells as well as move one cell to another.

We say that an arbitrary group element $g \in G$ *respects the set* A if either $g(A) = A$, $(g \in \mathcal{S}\mathrm{ym}(A))$, or $g(A) \cap A = \phi$ (i.e., A and $g(A)$ are disjoint). The symmetries $\mathcal{S}\mathrm{ym}([A; A_i])$ of a partition $[A; A_i]$ are the symmetries of the set A that respect all of the cells A_i of the partition. Such transformations are said to *respect the partition*. Note that for any symmetry g of the partition, two elements $a, b \in A$ belong to the same cell A_i iff $g(a), g(b) \in A$ belong to the same cell A_j.

The *stabilizer* of an object, w.r.t. G, is the subgroup of G that fixes the components of the object

$$\begin{aligned}\mathcal{S}\mathrm{tab}_G(a) &\equiv \{g \in G \mid g(a) = a\}\\ \mathcal{S}\mathrm{tab}_G(A) &\equiv \{g \in G \mid g(a) = a, \forall a \in A\}\\ \mathcal{S}\mathrm{tab}_G([A; A_i]) &\equiv \{g \in G \mid g(A_i) = A_i, \forall A_i \in [A; A_i]\}.\end{aligned}$$

Again, we write $\mathcal{S}\mathrm{tab}(\cdot)$ when G is clear from context. Equivalently

$$\begin{aligned}\mathcal{S}\mathrm{tab}(a) &= \mathcal{S}\mathrm{ym}(a)\\ \mathcal{S}\mathrm{tab}(A) &= \cap_{a \in A}\mathcal{S}\mathrm{ym}(a)\\ \mathcal{S}\mathrm{tab}([A; A_i]) &= \cap_{A_i \in [A; A_i]}\mathcal{S}\mathrm{ym}(A_i).\end{aligned}$$

The stabilizer of a set maps every point in the set to itself, while the stabilizer of a partition $[A; A_i]$ maps each cell of the partition onto itself, although the points within the cell may be permuted.

Note that the stabilizer of an object is always a normal subgroup of the symmetry group of the object and in particular

$$\mathcal{S}\mathrm{tab}([A; A_i]) \lhd \mathcal{S}\mathrm{ym}([A; A_i]).$$

In this case, we define the *partition permutation group* as the quotient group

$$\mathcal{P}\mathrm{pg}([A; A_i]) \equiv \mathcal{S}\mathrm{ym}([A; A_i])/\mathcal{S}\mathrm{tab}([A; A_i]).$$

This group represents the set of permutations of the cells under the symmetries of the partition.

Two sets $A, A' \subseteq S$ are said to be *congruent* (w.r.t. G) if there exists a transformation $g \in G$ such that $g(A) = A'$. Congruent sets have conjugate symmetry groups

$$\mathcal{S}\mathrm{ym}(A') = g\mathcal{S}\mathrm{ym}(A)g^{-1}.$$

If G is a group of isometries of a metric space S, then the sets A and A' are said to be *geometrically congruent*. Similarly, let $[A; A_i]$ and $[A; A_i']$ be two partitions of the same set $A \subseteq S$. Suppose a map $g \in G$ has the property that $g(a)$ and $g(b)$ lie in the same cell of the partition $[A; A_i']$ iff a and b lie in the same cell of the partition $[A; A_i]$. Then the transformation $g \in G$ of the set S induces an invertible transformation from the partition $[A; A_i]$ to the partition $[A; A_i']$. We then say the two partitions are congruent. If two partitions are congruent, then their symmetry groups are conjugates of one another

$$\mathcal{S}\mathrm{ym}([A; A_i']) = g\mathcal{S}\mathrm{ym}([A; A_i])g^{-1}.$$

B. Group Actions on Blocks

As we will see, the problem of complete characterization of group code partitions is very closely related to the concept of *group actions on blocks* [9], a topic from classical group theory which we briefly introduce here.

Let A be a nonempty subset of a set $A \subseteq S$, and let G be a subgroup of the symmetry group $G < \mathcal{S}\mathrm{ym}(A)$ that is transitive on A. A subset $B \subseteq A$ is said to be a *block* of G if for any $g \in G$ either the image is contained in B, $g(B) \subseteq B$, or is disjoint from B, $g(B) \cap B = \phi$. Clearly, the entire set A, the empty set ϕ, and the singleton sets $\{a\}$, where $a \in A$, are blocks of G; these are regarded as the *trivial* blocks. Here are some of the simple properties of blocks.

Fact 6: If a subset $B \subseteq S$ is a block, then for any $g \in G$, the following are equivalent: $g(B) \cap B \neq \phi$, $g(B) \subseteq B$, and $g(B) = B$.

Thus if $g \in G$ satisfies any of these conditions, then $g \in \mathcal{S}\mathrm{ym}(B)$. If a subset $B \subseteq S$ is a block of G then every element of $g \in G$ respects B. □

Fact 7: If B is a block and $H = \mathcal{S}\mathrm{ym}(B)$, then for any $a \in B$, we have $H(a) = B$.

Thus a block B is a group code. □

Fact 8: If $a \in A$ and $H < G$, then $H(a)$ is a block iff $H \cdot \mathcal{S}\mathrm{tab}(a) \cdot H \subseteq H \cdot \mathcal{S}\mathrm{tab}(a)$ (where $H \cdot G \equiv \{hg \mid h \in H, g \in G\}$) In this case, we have

$$\mathcal{S}\mathrm{ym}(H(a)) = H \cdot \mathcal{S}\mathrm{tab}(a) \cdot H.$$

This is because

$$g(H(a)) \cap H(a) \neq \phi \text{ iff } g \in H \cdot \mathcal{S}\mathrm{tab}(a) \cdot H$$

and for all $g \in H \cdot \mathcal{S}\mathrm{tab}(a) \cdot H$

$$gH(a) \subseteq H(a) \text{ iff } g \in H \cdot \mathcal{S}\mathrm{tab}(a).$$

Note that this condition is satisfied whenever $H \lhd G$ or $\mathcal{S}\mathrm{tab}(a) < H$. □

Fact 9: Conjugate groups induce congruent blocks.

Let G_1 and G_2 be transitive subgroups of G, and let $G_2 = gG_1g^{-1}$. Then, B is a block of G_1 with symmetry group $H = \mathcal{S}\mathrm{ym}(B)$ iff $B' = g(B)$ is a block of G_2 with symmetry group $H' = \mathcal{S}\mathrm{ym}(B') = gHg^{-1}$. □

Fact 10: If B is a block of G, then the distinct sets $A_i = g_i(B)$, where $g_i \in G$, are disjoint, and constitute a partition $[A; A_i]$ of the set $A = \cup_i A_i = G(B)$.

Every cell of the above partition is itself a block of G and together, they constitute a *complete block system* of G associated with the block B. □

Fact 11: A partition $[A; A_i]$ of A is a complete block system of *some* subgroup of G iff the symmetry group $\mathcal{S}\mathrm{ym}_G([A; A_i])$ is transitive on A.

The symmetry group $\mathcal{S}\mathrm{ym}_G([A; A_i])$ is the unique, maximal subgroup of G with respect to which the partition is a complete block system. The partition $[A; A_i]$ is a complete block system of G iff every transformation $g \in G$ respects the partition, i.e., $G = \mathcal{S}\mathrm{ym}_G([A; A_i])$. □

Fact 12: Conjugate subgroups lead to congruent complete block systems.

If G_1 and G_2 are conjugate subgroups of G that are transitive on $A \subseteq S$, with $G_1 = g^{-1}G_2g$, then the complete block system of G_1 associated with any block $B \subseteq A$ is congruent to the complete block system of G_2 associated with the block $g(B)$. □

Finally, we have the following classical result from the theory of group actions on blocks.

Theorem 13 [9, p. 14, Theorem 7.4]: The lattice of subgroups between $\mathcal{S}\mathrm{tab}(a), a \in A \subseteq S$, and G is isomorphic to the lattice of all blocks of the set A that contain the point a.

Proof of Theorem 13: Associate any block B of G containing $a \in A$ with the subgroup $H = \mathcal{S}\mathrm{ym}_G(B)$, which evidently contains $\mathcal{S}\mathrm{tab}(a)$, and conversely, associate every subgroup $H < G$ containing $\mathcal{S}\mathrm{tab}(a)$ with the $H(a)$, which is a block by virtue of Fact 8. The fact that this correspondence is unique follows from the observations $\mathcal{S}\mathrm{ym}(B)(a) = B$ (see Fact 7) and $\mathcal{S}\mathrm{ym}(H(a)) = H$ (Fact 8), which holds for all blocks B containing a and all subgroups H containing $\mathcal{S}tab(a)$. □

Corollary 14: There exists a subgroup H such that $\mathcal{S}\mathrm{tab}(a) < H < G$ (proper inclusions) iff there exists a block B such that $\{a\} \subseteq B \subseteq A$ (proper inclusions). □

C. Group Code Partitions and Geometrically Uniform Partitions

When designing GU trellis codes, one is interested in partitioning a GU signal set into congruent subsets.

Let $\mathbb{C}_0 \subseteq \mathbb{X}^{\mathsf{T}}$ be a group code over $(\mathbb{X}, \Sigma)$. Let $\Lambda < \Sigma$ be a subgroup of Σ that both respects the set $\mathbb{C}_0$ (i.e., every $\lambda \in \Lambda$ respects $\mathbb{C}_0$; the set $\lambda(\mathbb{C}_0)$ is either equal to $\mathbb{C}_0$ or disjoint from $\mathbb{C}_0$) and is transitive on $\mathbb{C}_0$. Then the group code defined by $\mathbb{C} \equiv \Lambda(\mathbb{C}_0)$ is said to be a *group code extension* of the subcode $\mathbb{C}_0$. If $\Gamma < \Lambda$ is a defining group of the group code $\mathbb{C}_0$, then the distinct sets determined by $\lambda_i(\mathbb{C}_0) \equiv \mathbb{C}_i$ are mutually disjoint, and constitute a partition of the group code $\mathbb{C}$. If x_0 is any point in the subcode $\mathbb{C}_0$, then we have $\mathbb{C}_0 = \Gamma(x_0)$ and

$$\mathbb{C}_i = \lambda_i(\mathbb{C}_0) = \lambda_i\Gamma\lambda_i^{-1}(\lambda_i(x_0))$$

so that the codes $\mathbb{C}_i$ form a collection of mutually congruent group codes. The resulting partition $[\mathbb{C}; \mathbb{C}_i]$ is said to be a *group code partition* of $\mathbb{C}$ w.r.t the subcode $\mathbb{C}_0$. If $\mathbb{X}^{\mathsf{T}}$ is a metric space and Σ is a group of isometries of $\mathbb{X}^{\mathsf{T}}$, then the codes $\mathbb{C}$ and $\mathbb{C}_0$ are GU codes. In this case, the code $\mathbb{C}$ is said to be a *geometrically uniform extension* of the isometry group code $\mathbb{C}_0$, and the partition $[\mathbb{C}; \mathbb{C}_i]$ is said to be a *geometrically uniform partition* of the isometry group code $\mathbb{C}$ w.r.t the subcode $\mathbb{C}_0$.

Given a group code $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$ over $\mathbb{X}$, and any subgroup $\Gamma < \Lambda < \Sigma$, the subcode

$$\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0) \subseteq \Lambda(\boldsymbol{x}_0)$$

is itself a group code. The orbits of the seed $\boldsymbol{x}_0$ under the action of the right cosets of Γ in Λ yield sets of group codes

$$\Gamma\lambda_i(\boldsymbol{x}_0) = \Gamma(\lambda_i(\boldsymbol{x}_0))$$

that have the same defining group Γ, and hence form a partition of $\Lambda(\boldsymbol{x}_0)$. In this case, the sets are *disjoint* but *may not be congruent*. On the other hand, the orbits of the seed $\boldsymbol{x}_0$ under the action of left cosets of Γ yields sets of congruent group codes

$$\mathbb{C}_i = \lambda_i(\mathbb{C}_0) = \lambda_i\Gamma\lambda_i^{-1}(\lambda_i(\boldsymbol{x}_0))$$

each with conjugate defining groups $\lambda_i\Gamma\lambda_i^{-1}$. Although the sets are *congruent* and the union of these group codes is the code $\mathbb{C}$, the subcodes $\mathbb{C}_i$ are *not necessarily disjoint*. In summary, given a subgroup $\Gamma < \Lambda$, the right cosets $[\Lambda; \Gamma\lambda_i]$ partition into subcodes that need not be congruent while the left cosets $[\Lambda; \lambda_i\Gamma]$ yield congruent subcodes that may not partition.

The key to obtaining a group code partition is the identification of subgroups $\Gamma < \Lambda$ that produce subcodes via the left cosets that are disjoint. Note that a sufficient (but not necessary) condition is that $\Gamma \lhd \Lambda$ is a normal subgroup. In this case, the left and right cosets agree and thus form a partition of congruent subcodes [2]. Another sufficient condition is that the stabilizer of the point $\boldsymbol{x}_0$ be a subgroup $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) < \Gamma$ [16]. Both necessary and sufficient conditions are presented below.

Consider a subgroup $\Gamma < \Lambda < \Sigma$ and a seed $\boldsymbol{x}_0 \in \mathbb{X}^{\mathsf{T}}$ and the group codes $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$, $\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0)$, and $\mathbb{C}_i = \lambda_i(\mathbb{C}_0)$. (The $\mathbb{C}_i$ are congruent, their union is $\mathbb{C}$, but they need not be disjoint.) The theory of blocks yields the following result regarding group partitions of the code $\mathbb{C}$ w.r.t $\mathbb{C}_0$ (i.e., determines when the $\mathbb{C}_i$'s are disjoint).

Theorem 15: The following statements are equivalent:

i) The subgroup $\Gamma < \Lambda$ defines a group code partition via the left cosets, i.e.

$$[\mathbb{C}; \mathbb{C}_i] = [\Lambda(\boldsymbol{x}_0); \lambda_i\Gamma(\boldsymbol{x}_0)].$$

ii) The subset $\mathbb{C}_0 \subseteq \mathbb{C}$ is a block of the group Λ.
iii)

$$\Gamma \cdot \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) \cdot \Gamma \subseteq \Gamma \cdot \mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0).$$

iv) $[\mathbb{C}; \mathbb{C}_i]$ is a complete block system of Λ.

Proof of Theorem 15: The equivalence of i) and ii) is clear from the definition of a block and a group code partition. The statements iii) and iv) follow from the theory of blocks.□

Note that the concept of group code extension described above is a natural generalization of our original definition of a group code, given in terms of a defining group and a seed. In the context of group code extensions, the subcode $\mathbb{C}_0 \subset \mathbb{X}^{\mathsf{T}}$ plays the role of the seed $x_0 \in \mathbb{X}^{\mathsf{T}}$ (a group code is a group code extension of a singleton set $\mathbb{C}_0 = \{\boldsymbol{x}_0\}$). Hence, we have the following properties of group code extensions, which are analogous to the properties of group codes, listed in Section II-D. In the foregoing, take $x_0 \in \mathbb{X}^{\mathsf{T}}$, $\Gamma < \Lambda$, $\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0)$, $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$, and require that all the elements of Λ respect the set $\mathbb{C}_0$ (this is a trivial restriction for a group code since every element of Σ respects every point $\boldsymbol{x}_0 \in \mathbb{X}^{\mathsf{T}}$). Note that $[\mathbb{C}; \mathbb{C}_i]$, where $\mathbb{C}_i = \lambda_i(\mathbb{C}_0), \lambda_i \in \Lambda$ is a group code partition of the code $\mathbb{C}$ w.r.t the subcode $\mathbb{C}_0$ and that each element of Λ respects every cell of the partition.

Fact 16 (Fact 1 Revisited): The group code partition $[\mathbb{C}; \mathbb{C}_i]$ is invariant under Λ.

This follows since $\mathbb{C}_0$ is a block of Λ □.

Fact 17 (Fact 2 Revisited): Any cell in $[\mathbb{C}; \mathbb{C}_i]$ can be used as the seed.

If $\mathbb{C}_1 \in [\mathbb{C}; \mathbb{C}_i]$, then

$$\Lambda(\mathbb{C}_1) = \Lambda(\boldsymbol{x}_0) = \lambda_1 \Gamma \lambda_1^{-1}(\lambda_1(\boldsymbol{x}_0))$$

and $[\mathbb{C}; \mathbb{C}_i]$ is the same partition generated by the subgroup $\lambda_1 \Gamma \lambda_1^{-1} < \Lambda$ and the cell $\mathbb{C}_1 = \lambda_1(\mathbb{C}_0)$. Thus a group code partition $[\mathbb{C}; \mathbb{C}_i]$ is determined by its generating group Λ and any cell $\mathbb{C}_i \in [\mathbb{C}; \mathbb{C}_i]$. This implies that in a group code partition, all the cells are on an equal footing. □

Fact 18 (Fact 3 Revisited): A partition $[\mathbb{C}; \mathbb{C}_i]$ over $\mathbb{X}$ is a group code partition iff its symmetry group $\mathcal{S}\mathrm{ym}_\Sigma([\mathbb{C}; \mathbb{C}_i])$ is transitive on the cells of $[\mathbb{C}; \mathbb{C}_i]$.

If Λ is any subgroup of $\mathcal{S}\mathrm{ym}_\Sigma([\mathbb{C}; \mathbb{C}_i])$ that is transitive on the group code partition $[\mathbb{C}; \mathbb{C}_i]$, then $[\mathbb{C}; \mathbb{C}_i] = \Lambda(\mathbb{C}_0)$ for any $\mathbb{C}_0 \in [\mathbb{C}; \mathbb{C}_i]$. Conversely, if $[\mathbb{C}; \mathbb{C}_i]$ is generated by $\mathbb{C}_0$ and by $\Lambda < \Sigma$ that respects $\mathbb{C}_0$, then Λ is a subgroup of $\mathcal{S}\mathrm{ym}([\mathbb{C}; \mathbb{C}_i])$ and is transitive on $[\mathbb{C}; \mathbb{C}_i]$. □

Fact 19 (Fact 4 Revisited): The cardinality of a group code partition $[\mathbb{C}; \mathbb{C}_i]$ (i.e., the number of cells) divides the index of the subgroup $\Gamma < \Lambda$.

Two transformations $\lambda_1, \lambda_2 \in \Lambda$ map the seed $\mathbb{C}_0$ to the same cell $\mathbb{C}_i \in [\mathbb{C}; \mathbb{C}_i]$ iff they belong to the same left coset of the stabilizer group $\mathcal{S}\mathrm{tab}_\Lambda(\mathbb{C}_0)$. This establishes a one-to-one correspondence between the cells and the left cosets of $\mathcal{S}\mathrm{tab}_\Lambda(\mathbb{C}_0)$, hence the conclusion follows from Lagrange's Theorem. □

Now consider the case where $\mathbb{X}^{\mathsf{T}}$ is a metric space. Fix $\Gamma < \Lambda$, $\mathbb{C} = \Lambda(\boldsymbol{x}_0)$, and $\mathbb{C}_0 = \Gamma(\boldsymbol{x}_0)$ where $[\mathbb{C}; \mathbb{C}_i]$ is a GU partition (i.e., a group code partition in a metric space). When the defining subgroup $\Gamma \lhd \Lambda < \mathcal{I}$so of the subcode $\mathbb{C}_0$ is normal, then the left and right coset partitions are the same and we have a GU partition of the isometry group code $\Lambda(\boldsymbol{x}_0)$. This special case is considered in [2]. A GU partition is also obtained when ever the subgroup Γ contains the stabilizer of the seed $\boldsymbol{x}_0$, $\mathcal{S}\mathrm{tab}_\Lambda(\boldsymbol{x}_0) < \Gamma$ [16]. The GU partition of a code w.r.t. a subcode need not be unique, as shown by Trott [16], who gave an example of a subcode that induces two (noncongruent) GU partitions (see Example 4).

We now use the theory of blocks to provide an answer to questions regarding all possible GU partitions. Given a subcode $\mathbb{C}_0$ of a group code $\mathbb{C} \subseteq \mathbb{X}^{\mathsf{T}}$, let $\mathcal{R}_{\mathbb{C}_0} \subset \mathcal{S}\mathrm{ym}(\mathbb{C})$ denote the set of all isometries that respect $\mathbb{C}_0$ (if $\lambda \in \mathcal{R}_{\mathbb{C}_0}$, then $\lambda(\mathbb{C}_0) = \mathbb{C}_0$, or $\lambda(\mathbb{C}_0) \cap \mathbb{C}_0 = \phi$). Note that, in general, $\mathcal{R}_{\mathbb{C}_0}$ is *not* a subgroup and that as a subset $\mathcal{S}\mathrm{ym}(\mathbb{C}_0) \subseteq \mathcal{R}_{\mathbb{C}_0}$. We say that two groups $\Lambda_1, \Lambda_2 \subseteq \mathcal{R}_{\mathbb{C}_0}$ are $\mathbb{C}_0$-*conjugate* if they are related by a conjugacy that fixes the subcode $\mathbb{C}_0$ (i.e., $\Lambda_2 = \lambda \Lambda_1 \lambda^{-1}$ for some $\lambda \in \mathcal{S}\mathrm{ym}(\mathbb{C}_0)$). The following theorem brings out the relationship between GU partitions of $\mathbb{C}$ and transitive subgroups of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ contained in the set $\mathcal{R}_{\mathbb{C}_0}$. Note that in the following, the subcode $\mathbb{C}_0$ is fixed, while the defining group, Λ, of $\mathbb{C}$ varies among the transitive subgroups of $\mathcal{S}\mathrm{ym}(\mathbb{C})$.

Theorem 20:

i) There exists a GU partition of a code $\mathbb{C}$ w.r.t. a subcode $\mathbb{C}_0$ iff the set $\mathcal{R}_{\mathbb{C}_0}$ contains a transitive subgroup of $\mathcal{S}\mathrm{ym}(\mathbb{C})$.
ii) The GU partition of the code $\mathbb{C}$ w.r.t. $\mathbb{C}_0$ is unique iff the set $\mathcal{R}_{\mathbb{C}_0}$ contains a *unique* maximal,[5] transitive subgroup of $\mathcal{S}\mathrm{ym}(\mathbb{C})$.
iii) The number of GU partitions of the code $\mathbb{C}$ w.r.t. $\mathbb{C}_0$ is equal to the number of maximal transitive subgroups of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ contained in the set $\mathcal{R}_{\mathbb{C}_0}$.
iv) The number of noncongruent GU partitions is equal to the number of non- $\mathbb{C}_0$-conjugate, maximal transitive groups contained in $\mathcal{R}_{\mathbb{C}_0}$.

Proof of Theorem 20: If Λ is any maximal transitive group contained in $\mathcal{R}_{\mathbb{C}_0}$, then $\mathbb{C}_0$ is a block of Λ, so it induces a complete block system $[\mathbb{C}; \mathbb{C}_i]$. Conversely, if $[\mathbb{C}; \mathbb{C}_i]$ is a complete block system with respect to $\mathbb{C}_0$, then the symmetry group $\Lambda = \mathcal{S}\mathrm{ym}([\mathbb{C}; \mathbb{C}_i])$ of the partition is a maximal transitive group contained in the set $\mathcal{R}_{\mathbb{C}_0}$. Moreover, any group $\Lambda' < \mathcal{S}\mathrm{ym}(\mathbb{C})$ that induces the same complete block system is necessarily a subgroup of Λ. This yields a one-to-one correspondence between the maximal transitive subgroups of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ contained in $\mathcal{R}_{\mathbb{C}_0}$ and distinct complete block systems associated with $\mathbb{C}_0$, or equivalently, distinct GU partitions of $\mathbb{C}$ w.r.t. $\mathbb{C}_0$. This proves assertions i), ii), and iii).

To establish the last assertion, if Λ_1 and Λ_2 are maximal groups in $\mathcal{R}_{\mathbb{C}_0}$ such that $\Lambda_1 = \lambda^{-1} \Lambda_2 \lambda$ for some $\lambda \in \mathcal{S}\mathrm{ym}(\mathbb{C}_0)$, then by Fact 11, the complete block system of Λ_1 associated with the block $\mathbb{C}_0$ is congruent to the complete block system of Λ_2 associated with the block $g(\mathbb{C}_0) = \mathbb{C}_0$. Conversely, let $\lambda \in \mathcal{S}\mathrm{ym}(\mathbb{C})$ define a congruence between two uniform partitions $[\mathbb{C}; \mathbb{C}_i]$ and $[\mathbb{C}; \mathbb{C}_i']$, and let $\lambda(\mathbb{C}_0) = \mathbb{C}_i'$. If $\lambda_2 \in \mathcal{S}\mathrm{ym}([\mathbb{C}; \mathbb{C}_i'])$ is a transformation such that $\lambda_2(\mathbb{C}_i') =$

[5]Maximal w.r.t. $\subseteq$ among subsets of $\mathcal{R}_{\mathbb{C}_0}$.

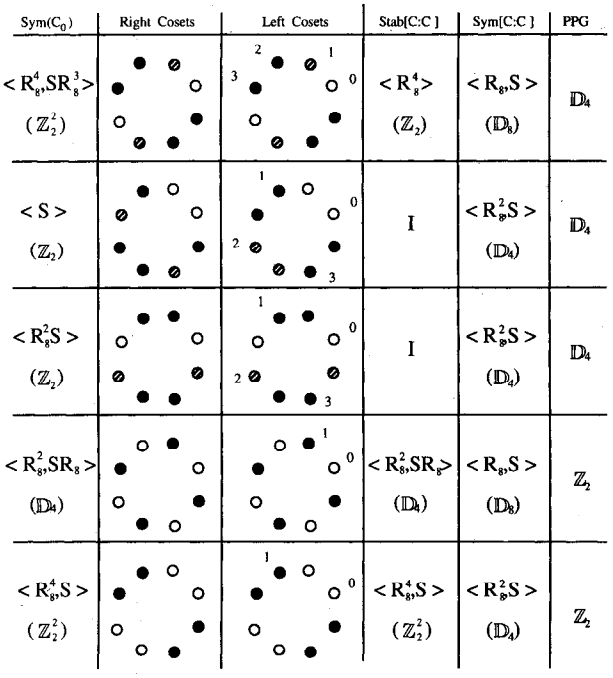

Fig. 5. The GU partitions of 8-PSK.

$\mathbb{C}_0$, then $\tilde{\lambda} \equiv \lambda_2; \lambda \in \mathcal{S}\mathrm{ym}(\mathbb{C}_0)$, and we have the conjugacy

$$\mathcal{S}\mathrm{ym}([\mathbb{C}; \mathbb{C}_i]) = \tilde{\lambda}^{-1}\mathcal{S}\mathrm{ym}([\mathbb{C}; \mathbb{C}_i'])\tilde{\lambda}. \qquad \square$$

This theorem provides a dual to the classical results in group theory presented in Theorem 13 and Corollary 14.

D. Examples of Geometrically Uniform Partitions

Example 3 (The GU Partitions of 8-PSK): Fig. 5 shows all GU partitions (up to congruence) of 8-PSK. For each partition, the symmetry group of the subcode $\mathbb{C}_0$ is given, followed by the stabilizer of the partition $\mathcal{S}\mathrm{tab}([\mathbb{C}; \mathbb{C}_i])$, the symmetry group of the partition $\mathcal{S}\mathrm{ym}(\mathbb{C}; \mathbb{C}_i])$, and the partition permutation group. Note that for the symmetry and stabilizer groups, the group is presented in terms of generators from $\mathcal{S}\mathrm{ym}(\mathbb{C}) \cong \mathbb{D}_8$, as well as a group isomorphic to it (e.g., $\langle S \rangle \cong \mathbb{Z}_2$). Additionally, for subgroups that are not normal in $\mathbb{D}_8$, the right coset partition is also shown.

Example 4 (The GU Partitions of the Cube): Consider the code $\mathbb{C}$ of the vertices of a cube in $\mathbb{R}^3$, with the permutation labeling shown in Fig. 6(a). Trott demonstrated that the subcode [1], [2] induces two different partitions [16], shown here in Fig. 6(b) and (c). This can be accounted for by considering the symmetry group $\mathcal{S}\mathrm{ym}(\mathbb{C})$. The group has 48 elements, and the subgroup lattice of $\mathcal{S}\mathrm{ym}(\mathbb{C})$ consists of 98 subgroups. Potential GU partitions can be considered from the subcodes generated by the intransitive subgroups acting on the vertex labeled 1, whose length divides $|\mathbb{C}| = 8$ (note that there exist subcodes of length 3 and 6, such as $[1, 3, 8]$). Of these subcodes, there are seven of length 2 and six of length 4. Up to congruence, however, we have the following six potential blocks: $[1, 2]$, $[1, 3]$, $[1, 7]$, $[1, 2, 3, 4]$, $[1, 2, 7, 8]$, and $[1, 3, 6, 8]$. There are 16 transitive subgroups of $\mathcal{S}\mathrm{ym}(\mathbb{C})$, and after applying Theorem 20 we see that the blocks $[1, 2]$ and $[1, 3]$ each induce two noncongruent GU partitions. All

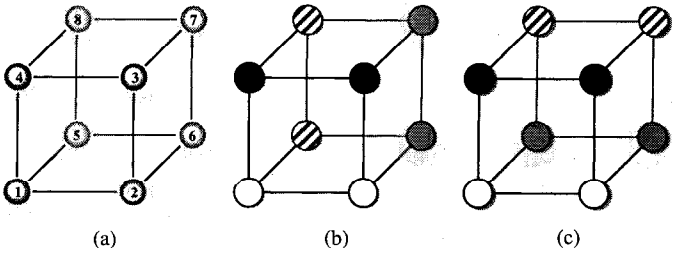

Fig. 6. The GU partitions of the cube.

TABLE I
THE PARTITION PERMUTATION GROUPS OF THE TWO-DIMENSIONAL UNGERBOECK PARTITIONS

Partition	$\mathcal{S}\mathrm{tab}([\mathbb{C};\ \mathbb{C}_i])$	$\mathcal{P}\mathrm{pg}$
2-way	$< R^2, S, {T_x}^2, {T_y}^2, T_xR, T_yR^3 >$	$\mathbb{Z}_2$
4-way	$< {T_x}^2, {T_y}^2, T_xR, T_yR^3 >$	$\mathbb{D}_4$
8-way	$< {T_x}^2{T_y}^2, {T_x}^2{T_y}^{-2} >$	$(\mathbb{Z}_2 \times \mathbb{Z}_4) \rtimes \mathbb{D}_4$
16-way	$< {T_x}^4, {T_y}^4 >$	${\mathbb{Z}_4}^2 \rtimes \mathbb{D}_4$

other subcodes considered are blocks of a unique element of the lattice of transitive subgroups. Hence, up to congruence, there are eight GU partitions of the cube

$$\{[1,2],[3,4],[5,8],[6,7]\}, \{[1,2],[3,4],[5,6],[7,8]\},$$
$$\{[1,3],[2,4],[5,7],[6,8]\}, \{[1,3],[2,5],[4,7],[6,8]\},$$
$$\{[1,7],[2,8],[3,5],[4,6]\}, \{[1,3,6,8],[2,4,5,7]\},$$
$$\{[1,2,3,4],[5,6,7,8]\}, \{[1,2,7,8],[3,4,5,6]\}. \qquad \square$$

Example 5 (The Ungerboeck Partitions in Two Dimensions): Consider the group of isometries of $\mathbb{R}^2$ described by $\Lambda = \langle R, S, T_x, T_y \rangle$, where R and S generate $\mathbb{D}_4$ (as before), while T_x and T_y are translations in the x and y directions. The QAM signal is generated by the action of Λ on the point $\boldsymbol{x}_0 = (\Delta/2, \Delta/2)$ (see Example 2).

The "standard" eight-way partition of the two-dimensional lattice translate can then be obtained by considering the subcode associated with the normal subgroup $\Gamma = \langle {T_x}^2{T_y}^2, {T_x}^2{T_y}^{-2} \rangle$. The subgroup Γ is also the stabilizer group $\mathcal{S}\mathrm{tab}([\mathbb{C}; \mathbb{C}_i])$ of the resulting GU partition. The group of transformations of this GU partition that is induced by the defining group Λ is isomorphic to the quotient group $\Lambda/\Gamma \cong (\mathbb{Z}_2 \times \mathbb{Z}_4) \rtimes \mathbb{D}_4$, which has 64 elements. In this semidirect product decomposition, the normal subgroup $\mathbb{D}_4$ corresponds to the group of rotations and reflections of the two–dimensional Euclidean space in the usual manner, whereas the subgroup $\mathbb{Z}_2 \times \mathbb{Z}_4$ corresponds to the translation group $\langle T_x, T_y \rangle$ modulo the stabilizer Γ. In particular, we may associate the element $(0, 1) \in \mathbb{Z}_2 \times \mathbb{Z}_4$ with the translation T_x modulo Γ and the element $(1, 0) \in \mathbb{Z}_2 \times \mathbb{Z}_4$ with the transformation $T_xT_yR_4^2$ modulo Γ. The partition permutation group has a transitive subgroup of 32 elements, given by the quotient group $\langle R_4, S, T_x^2, T_xT_yR_4^2 \rangle/\Gamma$. This partition and the above subgroup of 32 elements will be revisited in later examples. The symmetries of the 2-, 4-, and 16-way Ungerboeck partitions can be characterized similarly. The results are summarized in Table I.

E. Isometric Labelings

Forney introduced the basic idea of an *isometric labeling* [2], which we generalize here. Let $\mathbb{C} \subseteq \mathbb{X}^{\mathbb{T}}$ be any code over $\mathbb{X}$, and let Σ be a group of transformations of $\mathbb{X}^{\mathbb{T}}$. Let $\mathcal{L}$ be a finite set, and consider a many-to-one, onto map, $l\colon \mathbb{C} \to \mathcal{L}$, called the *labeling function*. That is, we assign a label $a_i \in \mathcal{L}$ to each point of the code $\mathbb{C}$.

An invertible transformation ω of the label set $\mathcal{L}$ is said to *respect* the labeling function l if there exists a map $\lambda \in \mathcal{S}\mathrm{ym}_{\Sigma}(\mathbb{C})$ such that $l(\lambda(x)) = \omega(l(x))$ for all $x \in \mathbb{C}$, i.e., we have the following commutative diagram:

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{l} & \mathcal{L} \\ \downarrow \lambda & & \downarrow \omega \\ \mathbb{C} & \xrightarrow{l} & \mathcal{L}. \end{array}$$

A group of transformations is said to respect the labeling function l if each of its elements does. If there exists a transitive group Ω of transformations of $\mathcal{L}$ that respects the labeling function l, then l is said to be a *group code labeling* of the code $\mathbb{C}$ with a *label group* Ω. The set of all invertible transformations of $\mathcal{L}$ that respect the labeling l constitutes a group, which we denote by $\mathcal{G}_l$. Clearly, a given labeling function l is a group code labeling w.r.t. *some* label group iff the group $\mathcal{G}_l$ is transitive on $\mathcal{L}$. In this case, any transitive subgroup of $\mathcal{G}_l$ may be chosen as the label group Ω.

Every labeling function $l\colon \mathbb{C} \to \mathcal{L}$ naturally induces a finite partition of $\mathbb{C}$, and conversely. Every transformation in the symmetries of the partition, $\lambda \in \mathcal{S}\mathrm{ym}_{\Sigma}([\mathbb{C}; \mathbb{C}_i])$, induces a map ω on the label set $\mathcal{L}$ that respects the labeling function l, and conversely (as seen by the commutative diagram). Moreover, two maps

$$\lambda_1, \lambda_2 \in \mathcal{S}\mathrm{ym}_{\Sigma}([\mathbb{C}; \mathbb{C}_i])$$

induce the same mapping $\omega \in \mathcal{G}_l$ iff λ_1 and λ_2 belong to the same coset of the stabilizer $\mathcal{S}\mathrm{tab}_{\Sigma}([\mathbb{C}; \mathbb{C}_i])$. Thus we see that the group $\mathcal{G}_l$ is isomorphic to the partition permutation group

$$\mathcal{P}\mathrm{pg}([\mathbb{C}; \mathbb{C}_i]) = \mathcal{S}\mathrm{ym}_{\Sigma}([\mathbb{C}; \mathbb{C}_i])/\mathcal{S}\mathrm{tab}_{\Sigma}([\mathbb{C}; \mathbb{C}_i]).$$

If $\mathbb{C}$ is a group code and the induced partition $[\mathbb{C}; \mathbb{C}_i]$ is a group code partition, then

$$\mathcal{P}\mathrm{pg}([\mathbb{C}; \mathbb{C}_i]) \ (\cong \mathcal{G}_l)$$

is transitive on the cells which implies that the labeling l is a group code labeling of $\mathbb{C}$.

If $\mathbb{X}^{\mathbb{T}}$ is a metric space and Σ is a group of isometries of $\mathbb{X}^{\mathbb{T}}$, a group code labeling $l\colon \mathbb{C} \to \mathcal{L}$ is said to be an *isometric labeling* of $\mathbb{C}$. Clearly, a labeling l of $\mathbb{C}$ is an isometric labeling of $\mathbb{C}$ if the associated partition is a GU partition. In this context, the group $\mathcal{G}_l$ is called the *label isometry group*.

A labeling function $l : \mathbb{C} \to \mathcal{L}$, together with the metric d on the set $\mathbb{X}$, induces a metric δ on the label set $\mathcal{L}$, i.e., $\forall a, b \in \mathcal{L}$

$$\delta(a, b) \equiv \min_{x,y \in \mathbb{C}, l(x)=a, l(y)=b} d(x, y).$$

Clearly, the induced metric is invariant under any map $\omega \in \mathcal{G}_l$; i.e., for every pair of labels $a, b \in \mathcal{L}$, we have $\delta(\omega(a), \omega(b)) = \delta(a, b)$.

When the label function determines an isometric labeling, the induced metric on $\mathcal{L}$ may be derived from a *weight function*, $w\colon \mathcal{L} \to \mathbb{R}^+$. Let $e \in \mathcal{L}$ be any fixed element of the label set $\mathcal{L}$. Since the label group $\Omega < \mathcal{G}_l$ is transitive, it is possible to choose a transformation $\omega_b \in \Omega$ for each $b \in \mathcal{L}$ such that $\omega_b(b) = e$. But then

$$\delta(a, b) = \delta(\omega_b(a), \omega_b(b)) = \delta(\omega_b(a), e).$$

It follows that if we define the weight function $w : \mathcal{L} \to \mathbb{R}^+$ on $\mathcal{L}$ by $w(a) \equiv \delta(a, e)$, then we have $\delta(a, b) = w(\omega_b(a))$.

Given a labeling function for a code in a metric space $\mathbb{X}^{\mathbb{T}}$, a permutation π of the labels that preserves the induced metric, $\delta(\pi(a), \pi(b)) = \delta(a, b)$, is called a *metric-preserving transformation* of the labels. The set of all such transformations forms a group $\mathcal{F}_l$, the *metric-preserving group* of the labeling. For an isometric labeling, the metric-preserving group contains the label isometry group $\mathcal{G}_l < \mathcal{F}_l$ as a subgroup, and in many cases, is strictly larger than $\mathcal{G}_l$. Note that in this case there are permutations of the cells that preserves the distances between the cells but are not obtained from isometries of $\mathbb{X}^{\mathbb{T}}$. These are equivalent to the so-called *distance profile-invariant transformations* described in [25].

As any label group Ω is required to be transitive, its cardinality is no smaller than that of the label set $\mathcal{L}$, and is, in general, strictly larger than $|\mathcal{L}|$. Nevertheless, the label isometry group $\mathcal{G}_l$ often has a *sharply transitive* subgroup Ω, i.e., a transitive subgroup $\Omega < \mathcal{G}_l$ such that $|\Omega| = |\mathcal{L}|$. In this case, the label set $\mathcal{L}$ may be identified with the label group Ω, so that the elements of the label group act on the labels by (left) translation, $h(a) = h*a$, $h \in \Omega$, $a \in \mathcal{L} = \Omega$. This corresponds to the situation considered in [2], where an isometric labeling was defined in terms of a label set $\mathcal{L}$ that was itself an (abelian) group. The distinctive feature of our expanded definitionis that the label group may have more elements than the size of the label set $|\mathcal{L}|$, and need not be an abelian group in general. An often beneficial form of isometric labeling is when the label set forms a ring (or a module over a ring) and the label group is a set of invertible, affine maps

$$\Omega = \{g \mid g_{a,b}(x) \equiv a \cdot x + b, a \in \mathcal{U}, b \in \mathcal{L}\}$$

where $\mathcal{U} \subset \mathcal{L}$ is the set of units (i.e., the invertible multiplicative elements) of the ring (or $\mathcal{U}$ is a set of invertible linear transformations on the module). (See Examples 6 and 7.)

Finally we note that our definition of isometric labeling is related to Loeliger's notion of a matched labeling [26], where he first considered many-to-one mappings from the symmetry group to a signal set. However, our definition differs in that we introduce an explicit label set $\mathcal{L}$. This explicit labeling is both a convenience for building encoders, as well as a necessity for certain classes of GU trellis codes whose system of symmetries are described in terms of a label group Ω, with $|\Omega| > |\mathcal{L}|$, as in Example 7 below.

F. Examples of Isometric Labelings

Example 6 (Phase-Shift Keying): First, consider the QPSK signal set (i.e., four points uniformly spaced on a circle centered at the origin); the constellation points form a square

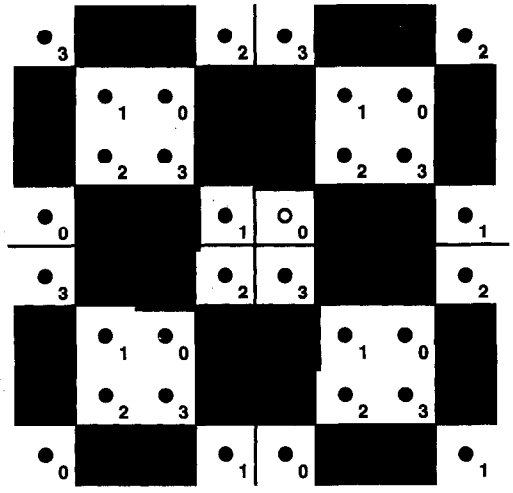

Fig. 7. An isometric labeling of QAM.

in $\mathbb{R}^2$. Suppose that all the four points are labeled distinctly. Then the label isometry group $\mathcal{G}_l$ coincides with the group of symmetries of the code (QPSK signal set). Thus $\mathcal{G}_l = \langle R_4, S\rangle \cong \mathbb{D}_4$. The label isometry group has two sharply transitive subgroups, namely, $\langle R_4\rangle \cong \mathbb{Z}_4$, and $\langle R_4^2, S\rangle \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$. We may therefore identity both Ω and $\mathcal{L}$ with any of these subgroups, and let the elements of Ω act on $\mathcal{L}$ by left translation. Alternatively, we may take the label isometry group $\mathcal{G}_l = \langle R_4, S\rangle = \langle R_4, R_4S\rangle$ and the label set $\mathcal{L} = \mathbb{Z}_4$, now regarded as a ring. The rotation R_4 acts on $\mathcal{L}$ as a translation $x \mapsto x + 1 \bmod 4$, while the reflection acts as a negation $x \mapsto -x \bmod 4$.

Now consider the 8-PSK signal set (i.e., eight points uniformly spaced on a circle centered at the origin), in which pairs of antipodal points have the same label. The corresponding label isometry group $\mathcal{G}_l$ may be identified with the quotient group $\langle R_8, S\rangle/\langle R_8^4\rangle \cong \mathbb{D}_4$. This leads to the same choice of label groups and group actions as the previous example. For isometric labelings of GU partitions, see Fig. 5. In both the above cases, the label isometry group $\mathcal{G}_l$ coincides with the metric-preserving group $\mathcal{F}_l$ of the labeling.

Example 7 (QAM): Now consider the infinite signal set (code) shown in Fig. 7. If one lets the points that are shaded be labeled by a "1" while those not shaded be "0" ("magnitude" label), together with the $\mathbb{Z}_4$ "phase" label of each point, one recognizes that this is an isometric labeling of the Ungerboeck eight-way partition of the two–dimensional lattice translate by $\mathbb{Z}_2 \times \mathbb{Z}_4$. Note that the windmill outlined in Fig. 7 can be taken as representatives of the cells of the partition.

If we consider only the phase labels of these points, the label isometry group corresponding to the labeling is seen to be identical to the symmetry group of the signal set, as before

$$\mathcal{G}_l = \mathcal{S}\mathrm{ym}(\mathbb{C}) = \langle R_4, S\rangle \cong \mathbb{D}_4.$$

On the other hand, if we also take the magnitude labels, then the label set has eight points, and the resulting label isometry group $\mathcal{G}_l$ is a group of 64 elements, isomorphic to the partition permutation group introduced in Example 5. That is, $\mathcal{G}_l = \Lambda/\Gamma$, where

$$\begin{aligned}\Lambda &= \langle R_4, S, T_x, T_y\rangle \\ &= \langle R_4, S, T_x, T_xT_yR_4^2, {T_x}^2{T_y}^2, T_x{T_y}^{-2}\rangle\end{aligned}$$

and

$$\Gamma = \langle {T_x}^2{T_y}^2, T_x{T_y}^{-2}\rangle.$$

The label isometry group has a transitive subgroup Ω of 32 elements, which corresponds to the quotient group

$$\langle R_4, S, {T_x}^2, T_xT_yR_4^2, {T_x}^2{T_y}^2, T_x{T_y}^{-2}\rangle/\ \Gamma.$$

If we now identify the label set $\mathcal{L}$ with the finite $\mathbb{Z}$-module $\mathbb{Z}_2 \times \mathbb{Z}_4$, then the elements of Ω act on it as invertible affine transformations. The elements of Ω may be expressed in the form

$$(A, b) : \mathbb{Z}_2 \oplus \mathbb{Z}_4 \to \mathbb{Z}_2 \oplus \mathbb{Z}_4, \boldsymbol{x} \mapsto A\boldsymbol{x} + b$$

where A is a 2×2 matrix, and b is an element of $\mathbb{Z}_2 \times \mathbb{Z}_4$. In this notation, we have

$$\begin{aligned} R_4 \bmod \Gamma &\equiv R : \boldsymbol{x} \mapsto \boldsymbol{x} + (0,1)^t \\ S \bmod \Gamma &\equiv S : \boldsymbol{x} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\boldsymbol{x} \\ T_xT_yR_4^2 \bmod \Gamma &\equiv T : \boldsymbol{x} \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\boldsymbol{x} \\ {T_x}^2 \bmod \Gamma &\equiv M : \boldsymbol{x} \mapsto \boldsymbol{x} + (1,0)^t. \end{aligned}$$

The group above may be defined in an abstract form in terms of its generators

$$\Omega = \langle R, S, T, M\rangle$$

together with the following relations:

$$\begin{aligned} &R^4 = S^2 = T^2 = M^2 = I, \\ &SR = R^3S, TR = RTM, MR = RM, \\ &TS = ST, MS = SM, MT = TM. \end{aligned}$$

We shall study this group again in the next section.

The isometric labeling shown above provides an example of the situation where the metric-preserving group $\mathcal{F}_l$ is strictly larger than the label isometry group $\mathcal{G}_l$. The permutation π on the label set $\mathcal{L}$ which swaps the elements $(0,1)$ and $(1,0)$ while leaving the other elements fixed is a metric-preserving transformation of the label set which is not obtained from any isometry of the Euclidean space. The metric-preserving group $\mathcal{F}_l$ corresponding to this labeling consists of 128 elements, and is generated by the label isometry group $\mathcal{G}_l$, together with the permutation π.

As mentioned earlier, the set of unshaded points in Fig. 7 is an example of a group code which is not a lattice. This signal set is a group code since it is generated as the orbit of the point

$$\boldsymbol{x}_0 = \left(\frac{1}{2}, \frac{1}{2}\right) \in \mathbb{R}^2$$

under the action of the group of isometries

$$\Lambda = \langle S, R, T_xT_y, T_x{T_y}^{-1}\rangle.$$

It is not a lattice since it is not closed under addition in $\mathbb{R}^2$. However, as it is GU, it is called a *regular array.*

IV. Classification of Symbolic Dynamic Groups and Orbit Systems

In this section we develop the concept of a *group system* (or a *symbolic dynamic group*). We then introduce the *orbit system of a symbolic dynamic group* as the image of a group system under its action on a "seed" sequence (that is typically over some label alphabet). We begin with an overview of *symbolic dynamics* and provide a framework under which we may classify group systems and their orbits within the context of symbolic dynamic systems.

Symbolic dynamics deals with collections of sequences satisfying certain constraints. Typically, the constraints are expressed in terms of a finite-state machine. We classify collections of sequences based on the properties of the finite-state machine used to generate them. We study the relationship between two such systems by means of *sliding block maps*, which will be described shortly. Two systems are considered to be equivalent or *conjugate*, if there exists an invertible sliding block map from one onto the other.

In this setup, a symbolic dynamic group is a set of sequences over a finite group, that is closed under component-wise group operation. It turns out that the finite-state machine generating a group system has a very special structure, and this yields a great deal of insight into the properties of group systems. An orbit of a symbolic dynamic group is obtained by regarding the elements of the finite group as invertible transformations on a set, and using this interpretation to generate sequences over the set from sequences over the finite group. This construction is a key step in the description of trellis group codes. It turns out that orbits of symbolic dynamic groups retain some but not all of the structure of group systems.

In this section, we will see that there are no nontrivial group systems over cyclic groups. Indeed, any (irreducible) group system over a cyclic group is a collection of all sequences over a certain alphabet. Such an unconstrained set of sequences is called a *fullshift*, and may be described by a trivial, one-state machine with several self-loops. Likewise, we will show that all (irreducible) group systems over dihedral groups may be decomposed into two components, one of which consists of unconstrained sequences generated by a one-state machine, while the other is a rate $1/2$ binary convolutional code with block length 2. As a consequence, we will see later that there are no interesting rotationally invariant codes over two-dimensional M-PSK signal set.

On a more positive note, we shall obtain useful structural characterization of orbits of symbolic dynamic groups. It is shown in [13] that group systems are characterized by two distinctive the properties: i) every irreducible group system has a unique, minimal finite-state machine generating it; ii) given an irreducible group system, there exists an invertible sliding block map from the group system to a set of unconstrained sequences (fullshift). The latter property is described by saying that every group system is *conjugate to a fullshift*. The most significant result that will be established in this section is that although orbits of symbolic dynamic groups may not have unique, minimal finite-state machines generating them, they are still conjugate to a fullshift. This means that it is possible to design an encoder from a set of unconstrained sequences to the given orbit system, together with a noncatastrophic decoder. The proof of the above result will also shed some light on the structure of orbit systems of symbolic dynamic groups.

A. A Review of Symbolic Dynamics

An *alphabet* $\mathcal{A}$ is a finite set, whose elements are referred to as *symbols* or *letters*. A *string*, *word*, or *block* over the alphabet $\mathcal{A}$ is an object of the form $a_1a_2a_3\cdots a_n$, obtained by concatenating *finitely* many symbols of the alphabet, where a symbol may be repeated if necessary. By concatenating infinitely many symbols of the alphabet (with repetitions), we obtain *bi-infinite sequences* such as $\cdots a_{-2}a_{-1}a_0a_1a_2\cdots$, *left semi-infinite* sequences such as $\cdots a_{-3}a_{-2}a_{-1}$, and *right semi-infinite* sequences of the form $a_0a_1a_2\cdots$ over the alphabet $\mathcal{A}$. Henceforth, we shall use the term *sequence* to denote either a bi-infinite or semi-infinite sequence, such as those described above.

A collection of strings over an alphabet $\mathcal{A}$ is called a *formal language* over $\mathcal{A}$. It is a subset of $\mathcal{A}^*$, the collection of all strings over $\mathcal{A}$ (including the empty string). A collection of (infinite) sequences over $\mathcal{A}$ is referred to as a *sequence space*. A sequence space is said to be bi-infinite, left-semi-infinite, or right-semi-infinite depending on the kind of sequences it contains. We denote the set of all bi-infinite sequences by $\mathcal{A}^{\mathbb{Z}}$, the set of right-semi-infinite sequences by $\mathcal{A}^{+\infty}$, and the set of left semi-infinite sequences by $\mathcal{A}^{-\infty}$.

We turn the sequence space $\mathcal{A}^{\mathbb{Z}}$ into a topological space by providing it with a topology whose open sets are generated by cylinder sets of the form $\{x \in \mathcal{A}^{\mathbb{Z}} \mid x_n = a\}$, one for each $n = 0, \pm 1, \pm 2, \cdots$, and $a \in \mathcal{A}$ [27], [28]. This is the standard topology inherited by the product space from the discrete topology on the set $\mathcal{A}$. (As a result of finiteness of the alphabet $\mathcal{A}$, the sequence space $\mathcal{A}^{\mathbb{Z}}$ turns out to be a compact, Hausdorff, separable, metrizable, totally disconnected topological space.) Under this topology, the shift map $\sigma, [\sigma(x)]_i = x_{i+1}$, is a homeomorphism of $\mathcal{A}^{\mathbb{Z}}$.

A *symbolic dynamic system* or a *subshift* over the alphabet $\mathcal{A}$ is defined to be a topologically closed, shift-invariant subset of $\mathcal{A}^{\mathbb{Z}}$. The subshift $\mathcal{A}^{\mathbb{Z}}$ itself is more commonly referred to as the *full* $|\mathcal{A}|$*-shift* or simply, a *fullshift*. A subset $\mathcal{S} \subset \mathcal{A}^{\mathbb{Z}}$ is topologically closed iff it contains all sequences $x \in \mathcal{A}^{\mathbb{Z}}$ which have the property that for any pair of integers $-\infty < m \leq n < \infty$, there is a bi-infinite sequence $y \in \mathcal{S}$ such that $x|_m^n = y|_m^n$. The concept of topological closure in symbolic dynamics corresponds to the notion of *completeness* [15], [18], [29] in dynamical systems theory. In symbolic dynamics, the condition of topological closure is imposed to rule out pathological shift-invariant subsets of $\mathcal{A}^{\mathbb{Z}}$. The following examples illustrate shift-invariant subsets that are not closed.

Example 8 (Topological Closure): Let $\mathcal{A} = \{0, 1\}$, the binary alphabet.

a) Consider the shift-invariant set

$$\mathcal{S} = \left\{x \in \mathcal{A}^{\mathbb{Z}} \mid \exists\, n \in \mathbb{Z} \text{ such that } x_i = \begin{cases} 0, & \text{if } i \leq n \\ 1, & \text{if } i > n \end{cases}\right\}.$$

The set $\mathcal{S}$ does not contain the all-zero sequence (or the all-one sequence), but every finite string of the all-zero sequence (and the all-ones sequence) is contained in some sequence in $\mathcal{S}$. Thus the topological closure of the $\mathcal{S}$ is obtained by adjoining these two sequences to it.

b) Alternatively, let $\mathcal{S}$ be the shift-invariant set consisting of all (bi-infinite) binary sequences with a finite and even number of ones

$$\mathcal{S} = \left\{ x \in \mathcal{A}^{\mathbb{Z}} \mid \sum_i x_i < \infty, \quad \sum_i x_i = 0 \bmod 2 \right\}.$$

Every finite string contained in every bi-infinite sequence over $\mathcal{A}$ is embedded in some sequence of $\mathcal{S}$. Hence, the closure of $\mathcal{S}$ is the entire fullshift over the binary alphabet. □

A subshift $\mathcal{S}$ is fully characterized by the set of all finite strings embedded in its bi-infinite sequences. This set of strings is called the *constraint set* of the subshift, denoted $\mathcal{C}(\mathcal{S})$. (We also use $\mathcal{C}_{-\infty}(\mathcal{S})$ and $\mathcal{C}_{+\infty}(\mathcal{S})$ to denote, respectively, the set of all left-semi-infinite and right-semi-infinite sequences embedded in bi-infinite sequences in the subshift $\mathcal{S}$.)

A function ϕ from a subshift $\mathcal{S}$ to a subshift $\mathcal{T}$ over the alphabets A and B, respectively, is said to be a *sliding block map* (or more specifically, an *n-block map*) if there exist integers $m, a > 0$ and a function $\tilde{\phi} : A^n \to B, n = m + a + 1$, such that for all $i \in \mathbb{Z}$ and all $x \in \mathcal{S}$, the ith component of $\phi(x)$ is given by $\tilde{\phi}(x_{i-m}x_{i-m+1} \cdots x_{i+a})$. Hedlund's theorem [30] asserts that sliding block maps are precisely the continuous, shift-commuting maps between the topological spaces $\mathcal{S}$ and $\mathcal{T}$. This implies that if a sliding block map has an inverse, then the inverse is also a sliding block map [10], [12], [28], [31], [32]. An invertible sliding block map is called a *conjugacy*, and two subshifts related by a conjugacy are said to be *topologically conjugate*.

Let $\mathcal{G}$ be a finite, directed, labeled graph, and let $\mathcal{S}$ be the set of label sequences generated by bi-infinite paths on this graph. The fact that the graph is finite ensures us that the set $\mathcal{S}$ is shift-invariant and topologically closed. Any subshift that may be described in terms of a finite graph in the above manner is referred to as a *sofic system* or *sofic shift*. For example, the subshift over the binary alphabet $\{0, 1\}$ which consists of all binary sequences in which zeroes are always separated by an even number of ones is a sofic system called the even shift. Fig. 8(c) depicts a directed, labeled graph that generates this subshift (where $a = 0$ and $b = 1$). The set of all binary sequences in which adjacent zeroes are separated by a prime number of ones is an example of a system which is *not* a sofic system.

If a finite, labeled graph $\mathcal{G}$ has the property that any bi-infinite label sequence is generated by at most one bi-infinite path on the graph, then the graph $\mathcal{G}$ is said to be a *conjugacy-inducing presentation* of the associated sofic system $\mathcal{S}$. If $\mathcal{G}$ is a conjugacy-inducing presentation of a sofic system $\mathcal{S}$, then there exists an invertible sliding block map from the set of all paths on the graph $\mathcal{G}$ onto the sofic system $\mathcal{S}$ [11], [14], [27], [28]. This means that the given sofic system is topologically conjugate to the set of all bi-infinite paths on the graph $\mathcal{G}$. A sofic system which has a conjugacy inducing presentation is called a *shift of finite type (SFT)*.

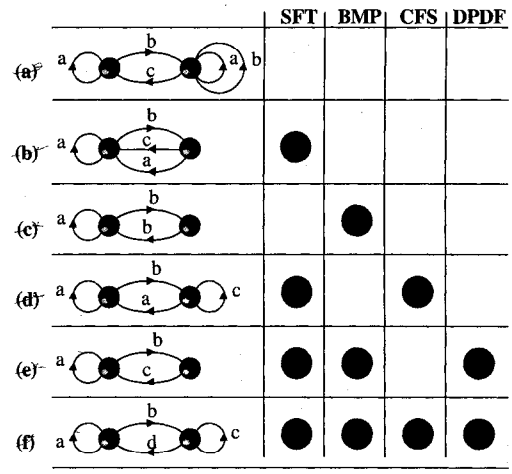

Fig. 8. Examples of irreducible sofic systems.

Shifts of finite type are in a sense the simplest and most useful class of subshifts, both from the point of view of symbolic dynamics and coding. A sofic system that is not an SFT is called a *proper sofic system*. The sofic systems in Fig. 8(b), (d), (e), and (f) are examples of SFT's, while those in Fig. 8(a) and (c) represent proper sofic systems. In the dynamical system terminology, a discrete-time system whose trajectories obey the SFT property are said to be *strongly complete* [15], [18], [17].

Let $\mathcal{S}$ be an arbitrary subshift (not necessarily a sofic system). If x is a left-semi-infinite sequence, we define the *future* of x to consist of all right-semi-infinite sequences y such that xy is a sequence in $\mathcal{S}$. More precisely

$$\mathbb{F}(x) = \{y \in \mathcal{A}^{+\infty} \mid xy \in \mathcal{S}\}, \qquad x \in \mathcal{A}^{-\infty}.$$

Similarly, the *past* of a right-semi-infinite sequence y may be defined as the set $\mathbb{P}(y)$ of all left semi-infinite sequences x such that wx is a sequence of $\mathcal{S}$

$$\mathbb{P}(y) = \{x \in \mathcal{A}^{-\infty} \mid xy \in \mathcal{S}\}, \qquad y \in \mathcal{A}^{+\infty}.$$

We can extend this notion by defining the future and past of finite strings, and the k-step future and k-step past of strings and semi-infinite sequences in the following manner:

$$\mathbb{F}(w) = \{y \in \mathcal{A}^{+\infty} \mid wy \in \mathcal{C}_{+\infty}(\mathcal{S})\}, w \in \mathcal{A}^*$$
$$\mathbb{P}(w) = \{x \in \mathcal{A}^{-\infty} \mid xw \in \mathcal{S}\}, w \in \mathcal{A}^*$$
$$\mathbb{F}_k(x) = \{y \in \mathcal{A}^k \mid xy \in \mathcal{C}(\mathcal{S}) \cup \mathcal{C}_{-\infty}(\mathcal{S})\}, x \in \mathcal{A}^* \cup \mathcal{A}^{-\infty}$$
$$\mathbb{P}_k(y) = \{x \in \mathcal{A}^k \mid xy \in \mathcal{C}(\mathcal{S}) \cup \mathcal{C}_{+\infty}(\mathcal{S})\}, y \in \mathcal{A}^* \cup \mathcal{A}^{+\infty}.$$

A sofic system may be characterized intrinsically, as a subshift whose (semi-infinite) sequences have only finitely many distinct futures, or equivalently, finitely many pasts [31], [27], [28]. That is, the sets

$$\{\mathbb{F}(x) \mid x \in \mathcal{C}_{-\infty}(\mathcal{S})\}$$

and

$$\{\mathbb{P}(y) \mid y \in \mathcal{C}_{+\infty}(\mathcal{S})\}$$

are finite. Similarly, an SFT is intrinsically characterized by the property that there exists an integer $N > 0$ such that for

any left semi-infinite sequence

$$\cdots x_{-3}x_{-2}x_{-1} \in \mathcal{C}_{-\infty}(\mathcal{S})$$

we have

$$\mathbb{F}(\cdots x_{-3}x_{-2}x_{-1}) = \mathbb{F}(x_{-N}\cdots x_{-2}x_{-1}).$$

This is equivalent to the condition that for any right-semi-infinite sequence $x_1x_2x_3\cdots \in \mathcal{C}_{+\infty}(\mathcal{S})$, we have

$$\mathbb{P}(x_1x_2x_3\cdots) = \mathbb{P}(x_1x_2\cdots x_N)$$

and amounts to saying that the subshift has a memory of a finite duration. From this definition it is clear that SFT's form a subclass of sofic systems. An SFT that satisfies the above condition with $N = n$ is called an n-step SFT.

Although the fact that a subshift has a finite number of distinct futures implies that it has a finite number of distinct pasts and *vice versa*, the cardinality of the two sets may be different. Consider, for example, the sofic system shown in Fig. 8(b). Since the system is a one-step SFT, the future of any left-semi-infinite sequence is determined by its last symbol, while the past of any right-semi-infinite sequence is determined by its first symbol. It is easily seen that this sofic system has two distinct futures, represented by $\mathbb{F}(a) = \mathbb{F}(c)$ and $\mathbb{F}(b)$, but three distinct pasts, given by $\mathbb{P}(a), \mathbb{P}(b)$, and $\mathbb{F}(c)$.

A labeled graph $\mathcal{G}$ is said to be *forward-deterministic* or just *deterministic* if no two outgoing edges from the same state have the same label. If all incoming edges into the same state have distinct labels, then the graph is said to be *backward-deterministic*. If a graph is both forward- and backward-deterministic, then it is said to be *bideterministic*. Any sofic system may be presented either by a backward-deterministic or a (forward-) deterministic graph, while only some sofic systems have bideterministic presentations. While the proper sofic system in Fig. 8(c) and the SFT's depicted in Fig. 8(e) and (f) clearly have bideterministic presentations, the proper sofic system in Fig. 8(a) and the SFT's in Fig. 8(b) and (d) do not admit any bideterministic presentations.

Given a labeled graph $\mathcal{G}$ presenting a sofic system $\mathcal{S}$, we define the future of a state s of $\mathcal{G}$ to be the set of all (right-semi-infinite) label sequences generated by paths originating at state s. The past of a state s is likewise defined to be the set of all label sequences generated by left-semi-infinite label sequences generated by paths terminating at state s. If two states of $\mathcal{G}$ have either the same past or the same future, then the two states may be *merged* together, without changing the sofic system represented by it. If a graph is forward- (backward-) deterministic, then the reduced graph obtained by merging all states with identical futures (pasts) is again forward- (backward-) deterministic. A labeled graph representing a sofic system is said to be *minimal* if no two states of the graph may be merged without changing the sofic system represented by it.[6] Every sofic system has a canonical forward-deterministic and backward-deterministic minimal presentation, known, respectively, as the *right Krieger cover* and *left Krieger cover* of the sofic system [33], [34].

If a subshift has the property that its left-semi-infinite sequences have *disjoint* futures, i.e., if for any two left-semi-infinite strings x and y, the futures $\mathbb{F}(x)$ and $\mathbb{F}(y)$ are either identical or disjoint, then we say that the subshift has the *disjoint future* property. Similarly, if the right-semi-infinite sequences of a subshift have disjoint pasts, then subshift is said to have the *disjoint past* property.

Two bi-infinite sequences x and y of a subshift $\mathcal{S}$ are said to be *past-equivalent* if

$$\mathbb{F}(x|_{-\infty}^{-1}) = \mathbb{F}(y|_{-\infty}^{-1})$$

and are said to be *future-equivalent* if

$$\mathbb{P}(x|_{0}^{\infty}) = \mathbb{P}(y|_{0}^{\infty}).$$

In general, sequences of a subshift could be past-equivalent without being future-equivalent, or *vice versa*. However

Proposition 21: The following statements about a subshift $\mathcal{S}$ are equivalent:

i) The subshift $\mathcal{S}$ has the disjoint future property.
ii) The subshift $\mathcal{S}$ has the disjoint past property.
iii) Two sequences of the subshift $\mathcal{S}$ are past-equivalent iff they are future-equivalent.

Proof of Proposition 21: Follows directly from the definitions. □

We call subshifts having disjoint pasts, or equivalently, disjoint futures, the *DPDF systems*. The SFT's in Fig. 8(e) and (f) are examples of DPDF systems. If a sofic system $\mathcal{S}$ is DPDF, then it turns out that the states of its right Krieger cover have disjoint futures and the states of its left Krieger cover have disjoint pasts. On the other hand, we will see later that the states of a labeled graph can have disjoint pasts as well as disjoint futures, but the sofic system represented by it is not a DPDF system.

A subshift $\mathcal{S}$ is said to be *irreducible* or *topologically transitive* if for any two strings w and y in the constraint set $\mathcal{C}(\mathcal{S})$, we have a string x such that the concatenation wxy is also a string in $\mathcal{C}(\mathcal{S})$. Indeed, a sofic system is irreducible if it is presented by an irreducible graph. Moreover, we have the following interesting and helpful result from [35].

Fact 22: If a labeled graph $\mathcal{G}$ generates an irreducible sofic system $\mathcal{S}$, then an irreducible component of $\mathcal{G}$ generates $\mathcal{S}$.

See [35, Lemma 1]. □

All the sofic systems in Fig. 8 are irreducible. The above notion of irreducibility is akin to the notion of *controllability* [15], [18], [29] in the theory of dynamical systems. We have the following well-known minimality result about irreducible deterministic presentations of irreducible sofic systems.

Fact 23: Every irreducible sofic system $\mathcal{S}$ has an irreducible, **deterministic** presentation $\mathcal{G}_0$ with the following properties:

i) The future of every state of $\mathcal{G}_0$ coincides with the future of some left semi-infinite sequence in $\mathcal{C}_{-\infty}(\mathcal{S})$.
ii) If $\mathcal{G}$ has an edge labeled a from state s to state t and x is any left semi-infinite sequence such that $\mathbb{F}(s) = \mathbb{F}(x)$, then $\mathbb{F}(t) = \mathbb{F}(xa)$.

[6] Note that our notion of minimality is based on the above "merger" property, and not on the number of states in the graph.

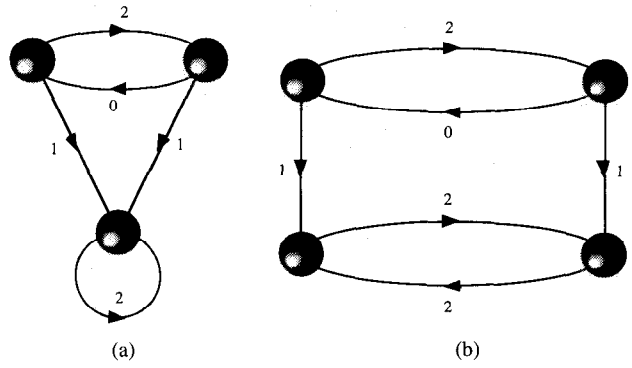

Fig. 9. A nonirreducible SFT with a nonminimal BP.

iii) $\mathcal{G}_0$ is a minimal presentation of $\mathcal{S}$.

iv) Given any other irreducible, deterministic graph $\mathcal{G}$ generating the sofic system $\mathcal{S}$, there exists a label-preserving graph homomorphism ϕ from $\mathcal{G}$ onto $\mathcal{G}_0$ which maps every state of $\mathcal{G}$ to the unique state of $\mathcal{G}_0$ that has the same future.

See [10], [12], [31]–[33]. □

As a consequence of the second conclusion above, if $\mathcal{S}$ is an n-step SFT, then every path on the graph generating a given label string of length n or more has a unique terminal state. Hence, if $\mathcal{S}$ is an SFT, then every (infinite) label sequence in the system is produced by a unique path on the graph $\mathcal{G}_0$.

The last conclusion in Fact 23 implies that for a given irreducible sofic system, the graph $\mathcal{G}_0$ is unique up to a label-preserving graph isomorphism. This canonical graph is called the *right Fischer cover* of their reducible sofic system $\mathcal{S}$ [33]. The right Fischer cover of an irreducible sofic system is a subgraph (in fact, an irreducible component) of the right Krieger cover. Similar statements are true with regard to backward-deterministic presentations and lead to the notion of the *left Fischer cover* of irreducible sofic systems.

From Fact 23, it follows that if either the left or the right Fischer cover of an irreducible sofic system turns out to be bideterministic, then the two Fischer covers are identical, and constitute the unique, irreducible, bideterministic minimal presentation of the given sofic system. An irreducible sofic system with the above property is said to be a *Bideterministic Minimal Presentation* system, or simply a BMP system. It is clear that an irreducible sofic system is a BMP system iff it has some minimal, bideterministic presentation (BP). In fact

Proposition 24: If an irreducible sofic system admits any bideterministic presentation, then it is a BMP system.

Proof of Proposition 24: See the Appendix. □

The proper sofic system in Fig. 8(c) and the SFT in Fig. 8(e) and (f) are BMP systems since the given representations are themselves bideterministic and minimal. As indicated before, the proper sofic system in Fig. 8(a) and the SFT's in Fig. 8(b) and (d) do not have such presentations. The nonirreducible SFT in Fig. 9(a) has a bideterministic presentation as shown in Fig. 9(b), but no bideterministic presentation of the system can be *minimal*. This brings out the significance of irreducibility in Proposition 24.

We now proceed to relate the DPDF property to the BMP property of irreducible sofic systems.

Proposition 25: An irreducible sofic system is a DPDF system iff it is an SFT and a BMP system.

Proof of Proposition 25: See Appendix. □

Note that if $\mathcal{C}$ is a DPDF system with a bideterministic minimal graph $\mathcal{G}$, then $\mathcal{G}$ is the **unique** minimal presentation of $\mathcal{G}$, and has the smallest number of states among all the graph presentations of $\mathcal{G}$ [34]. Thus DPDF systems exhibit a rather strong form of minimality.

The BMP system in Fig. 8(c) is not an SFT and hence not a DPDF system. The SFT's in Fig. 8(b) and (d) are not BMP systems and, consequently, they are not DPDF systems either. The irreducible SFT's in Fig. 8(e) and (f) are BMP systems and, hence, they are also DPDF systems.

Our proof of Proposition 25 (presented in the Appendix) makes use of the notion of Fischer cover of irreducible sofic systems. The proof may be carried out without the irreducibility assumption, by working with Krieger covers in place of Fischer covers. This leads to the following amplification of Proposition 25 (a result essentially contained in [34]).

Proposition 26: For a sofic system $\mathcal{S}$, the following statements are equivalent:

i) $\mathcal{S}$ is a DPDF system.

ii) $\mathcal{S}$ is an SFT and its right Krieger cover is bideterministic.

iii) $\mathcal{S}$ is an SFT and its left Krieger cover is bideterministic.

Proof of Proposition 26: See [34, Theorem 1]. □

Two sliding block maps $\pi_1 : S_1 \to T_1$ and $\pi_2 : S_2 \to T_2$ are said to be *equivalent* or *conjugate* if there exist topological conjugacies $\phi : S_1 \to S_2$ and $\theta : T_1 \to T_2$ such that $\theta \circ \pi_1 = \pi_2 \circ \phi$, that is, the following diagram commutes:

$$\begin{array}{ccc} S_1 & \xrightarrow{\pi_1} & T_1 \\ \phi \downarrow & & \downarrow \theta \\ S_2 & \xrightarrow{\pi_1} & T_2. \end{array}$$

Given any subshift $\mathcal{S}$, we define a subshift $\mathcal{S}^{[n]}$ which consists of sequences of n-blocks of the form

$$\cdots x|_{-2}^{n-3} x|_{-1}^{n-2} x|_{0}^{n-1} x|_{1}^{n} x|_{2}^{n+1} \cdots$$

for every sequence $x \in \mathcal{S}$. The shift $\mathcal{S}^{[n]}$ is said to be a *higher block system* of $\mathcal{S}$ [10], [12], [28], [31], and is topologically conjugate to the subshift $\mathcal{S}$. This is a very important construction in symbolic dynamics, and is used to show any SFT is conjugate to a one-step SFT, and that any sliding block map is equivalent to a 1-block map (a sliding block map with window size 1).

We say that a subshift is CFS if it is conjugate to a fullshift over some alphabet. Sofic systems with this property are important from the point of view of coding, where one deals with the problem of encoding arbitrary message sequences (fullshifts) into constrained sequences (subshifts) in such a way that transformation from one representation to another is accomplished by means of a simple, reliable scheme (a sliding block map). Clearly, if a subshift is CFS, then it is an irreducible SFT. However, the irreducible SFT's of Fig. 8(b) and (e) are not CFS. This follows from the fact that for an irreducible SFT to be CFS, it is necessary (but not sufficient) that its topological entropy [10] (the logarithm of the largest eigenvalue of the adjacency matrix of the minimal deterministic presentation) is the logarithm of a positive integer. The

SFT's in Fig. 8(d) and (f) are CFS because the mapping that takes label sequences to the corresponding state sequences on the graph establishes a conjugacy between the given SFT and the full 2-shift. These examples show that SFT's could be DPDF systems without being CFS and conversely.

B. Symbolic Dynamic Groups

Suppose the alphabet under consideration is a finite group G. Then the fullshift $G^{\mathbb{Z}}$ may be regarded as a group under a component-wise group operation, thereby furnishing it with a direct product group structure. A subshift $\mathcal{S} \subseteq G^{\mathbb{Z}}$ is said to be a *symbolic dynamic group* or a *group system* if $\mathcal{S}$ is a subgroup of the fullshift $G^{\mathbb{Z}}$ under the above group structure. It may be seen that the group operation is a *continuous* function in the topology of symbolic dynamic systems and that the shift map is a *group automorphism* of the group system under the direct product group structure.

As shown in [13], [17], [18], group systems exhibit some very remarkable properties. First of all, they possess a strong form of the DPDF property: For any two strings or sequences x and y embedded in sequences of a group system $\mathcal{S}$, if the concatenation

$$x|_{-m}^{-1} y|_0^n \quad (1 \leq m, n \leq \infty)$$

is a string or sequence embedded in a bi-infinite sequence of $\mathcal{S}$, then so is the concatenation

$$y|_{-m}^{-1} x|_0^n.$$

This means that any two k-step futures ($k > 0$) or infinite futures of two strings or (left)-semi-infinite sequences are either identical or disjoint. In fact, if e is the identity element of the group G, then the k-step future ($k > 0$) of the string e^n ($n > 0$) is a *normal subgroup* of[7] G^k. Furthermore, the infinite future of e^n is a normal subgroup of $G^{\mathbb{Z}+}$. In fact, the same is true of the k-step future and infinite future of the left-semi-infinite sequence $e^{\mathbb{Z}-}$ as well. The (finite or infinite) future of any other string or left-semi-infinite sequence is a *coset* of the normal subgroup corresponding to the future of the identity sequence of the same length. This leads to

Fact 27: All group systems are SFT's with the DPDF property.

The fact follows from the preceding statements on the future sets, together with the finiteness of the alphabet G [13]. □

The minimal graph $\mathcal{G}$ generating a group system $\mathcal{S}$ has the following structure. The states of the graph have the group structure given by $G^{\mathbb{Z}+}/\mathbb{F}(e^{\mathbb{Z}-})$. This means that whenever there is an edge labeled a_1 from state s_1 to state t_1 and an edge labeled a_2 from state s_2 to state t_2, there exists an edge labeled $a_1 * a_2$ from state $s_1 * s_2$ to the state $t_1 * t_2$. The set of labels $\mathcal{P}$ of the self-loops around the identity state is a normal subgroup of G, called the *parallel transition subgroup*. If there is an edge from a state s to a state t in the graph, then there are exactly $|\mathcal{P}|$ edges from s to t, whose labels constitute a coset of the parallel transition subgroup $\mathcal{P}$. The fullshift $\mathcal{P}^{\mathbb{Z}}$ is a group system that forms a normal subgroup of the system $\mathcal{S}$, and is called the *parallel transition shift*. The quotient group $\mathcal{S}/\mathcal{P}^{\mathbb{Z}}$ is a group system over the quotient group alphabet $G/\mathcal{P}$, which is referred to as the *quotient group shift*.

The set of labels of all outgoing edges of the identity state $F_1(e^{\mathbb{Z}-})$ is also a normal subgroup of G, called the *(forward) input group* [18]. The set of labels of all outgoing edges of any other state is a coset of this subgroup. We may similarly define the *backward input group* and make an analogous statement regarding its cosets, by focusing our attention on the incoming edges of the states in the graph. If the group system is a 1-step SFT, then all the edges of its minimal graph have distinct labels, so that the every state may be identified with the coset corresponding to its outgoing edges or incoming edges, which yields an isomorphism between the groups concerned.

From the characterization above of the minimal graph generating a group system $\mathcal{S}_1$ which is also a 1-step SFT, it follows that $\mathcal{S}_1$ is conjugate, by means of a one-block map with a one-block inverse, to the product of the group system $\mathcal{S}/\mathcal{P}^{\mathbb{Z}} \equiv \mathcal{S}_1'$ and the full $|\mathcal{P}|$-shift. The minimal graph generating $\mathcal{S}_1'$ is the same as that of $\mathcal{S}_1$, with the modification that each set of parallel edges is replaced by a single edge, labeled by the corresponding coset of $\mathcal{P}$. Since this graph has no nontrivial parallel edges and since all its edges are distinctly labeled, it follows that the group system $\mathcal{S}_1'$ is conjugate, by means of a 1-blockmap with a 2-block inverse, to the bi-infinite *state sequences* generated by the paths on the graph. But this subshift is readily identified with the group system $\mathcal{S}_1'/F^{\mathbb{Z}} \equiv \mathcal{S}_2$, where F is the input group of the group system $\mathcal{S}_1'$. Thus we see that the original group system $\mathcal{S}_1$ is conjugate to the product of the group system $\mathcal{S}_2$ and the full $|\mathcal{P}|$-shift. Unless the input group of $\mathcal{S}_1$ is trivial, the group system $\mathcal{S}_2$ has a strictly smaller alphabet than the group system $\mathcal{S}_1$. We may thus proceed by induction to show that the given group system $\mathcal{S}$ is conjugate to the product of several fullshifts and a group system $\mathcal{S}_n$ whose input group is trivial. If the group system $\mathcal{S}_1$ is irreducible, then so is $\mathcal{S}_n$, which implies that the minimal graph of $\mathcal{S}_n$ consists of a single state with a self-loop labeled by the identity element of the (group) alphabet. This shows

Fact 28: Every irreducible group system is conjugate to a fullshift, the size of whose alphabet is equal to the size of its input group.

The argument above is due to Kitchens and is presented in slightly greater generality in [13]. From the above results, it follows that irreducible symbolic dynamic groups lie in the intersection of DPDF and CFS systems and thus exhibit all the nice properties of subshifts defined in the previous section. □

The best known way of systematically enumerating group systems is by means of *cycles*, which is analogous to the characterization of convolutional codes in terms of generator matrices. A bi-infinite sequence over G is called a cycle if it is equal to the identity element at all but finitely many of its coordinates. Given a set of cycles, there is a unique minimal group system that contains them, called the group system *generated by the set of cycles*. Given a finite list of cycles, the minimal graph of the group system generated by these cycles may be obtained by using techniques discussed in [36]. Every irreducible group system is generated by some finite set of cycles, and hence may be realized in this manner.

[7] Here it is necessary to assume that every letter of the alphabet G occurs in some sequence of the subshift $\mathcal{S}$.

The characterization of irreducible group systems in terms of cycles has many other virtues. It has been used to show that the minimal graph of an irreducible group system has the structure of a *product of de Bruijn graphs* [18], [36]. We will presently use the cycle characterization to provide a complete classification of irreducible group systems over cyclic groups $\mathbb{Z}_n$ and dihedral groups $\mathbb{D}_n$. As we will see, these group systems are crucial in describing GU trellis codes over two-dimensional PSK signal sets.

Theorem 29: Every irreducible group system over a cyclic group $\mathbb{Z}_n$ is a fullshift.

Proof of Theorem 29: See the Appendix. □

This means that every irreducible group system over a cyclic group is a set of unconstrained sequences, generated by a trivial, one-state graph, with one or more parallel transitions. Recall that any group system may be decomposed into a parallel transition shift, representing a set of unconstrained sequences generated by a one-state graph, and a quotient group shift, which has a trivial parallel transition group. The following result shows that if the alphabet under consideration is a dihedral group, then the latter component is an irreducible group system over $\mathbb{Z}_2 \oplus \mathbb{Z}_2$, which is simply a rate $1/2$-binary convolutional code with block-length 2.

Theorem 30: The quotient group shift of an irreducible group system over a dihedral group is an irreducible group system over a group isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

Proof of Theorem 30: See the Appendix. □

C. The Orbit System of a Symbolic Dynamic Group

We shall now formally introduce the concept of an *orbit of a symbolic dynamic group*. Consider a finite group G that acts on a set X. If $\Lambda < G^{\mathbb{Z}}$ is a group system over G and x_0 is an element of the set X, we define a subshift over X by letting each component of the sequences in Λ act on the fixed element $x_0 \in X$. This subshift is denoted $\Lambda(x_0)$ and is said to be the *orbit system* generated by the group system Λ and the seed x_0. That is, $x \in \Lambda(x_0)$, where

$$x = \cdots (x_0)g_{-1}(x_0)g_0(x_0)g_1(x_0)\cdots \in X^{\mathbb{Z}}$$

for each $\cdots g_{-1}g_0g_1\cdots \in S$. This is a generalization of Slepian's notion of (finite) group codes and has proved to be useful in providing insight into the nature of certain "almost linear" codes [37].

We shall now provide a classification of trellis group codes along the same lines as we did for group systems. The main difference between a group system and an orbit of asymbolic dynamic group arises from the fact that the minimal graph corresponding to the orbit system may be strictly smaller than the minimal graph of the group system generating it. When this happens, the orbit system may or may not inherit some of the properties mentioned above from its parent subshift. We first consider an example of an orbit system whose minimal graph differs from that of the group system generating it, but which nevertheless continues to have all the properties of group systems above. We will then present an example in which the orbit system loses the DPDF and BMP properties, but continues to be CFS.

D. Examples of Orbit Systems

Both examples are based on the following group G of 32 elements, introduced in the previous section in connection with Ungerboeck eight-way partition of two-dimensional integer lattice. Recall that this group be may regarded as a set of affine transformations acting on the finite module $\mathbb{Z}_2 \oplus \mathbb{Z}_4$. The elements of $G = \langle R, S, T, M\rangle$ may be expressed in the form

$$(A, b) : \mathbb{Z}_2 \oplus \mathbb{Z}_4 \to \mathbb{Z}_2 \oplus \mathbb{Z}_4, \boldsymbol{x} \mapsto A\boldsymbol{x} + b$$

where A is a 2×2-matrix, and b is an element of $\mathbb{Z}_2 \times \mathbb{Z}_4$. In this notation, we have the following representation for the generators of G:

$$\begin{aligned} R &: \boldsymbol{x} \mapsto \boldsymbol{x} + (0,1)^t \\ S &: \boldsymbol{x} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\boldsymbol{x} \\ T &: \boldsymbol{x} \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\boldsymbol{x} \\ M &: \boldsymbol{x} \mapsto \boldsymbol{x} + (1,0)^t. \end{aligned}$$

The group may also be described in terms of its generators $G = \langle R, S, T, M\rangle$ and relations

$$\begin{aligned} &R^4 = S^2 = T^2 = M^2 = I, \\ &SR = R^3S, TR = RTM, MR = RM, \\ &TS = ST, MS = SM, MT = TM. \end{aligned}$$

Example 9 (An Orbit System with a Minimal BP): The smallest group system over the group alphabet above that contains the cycle of length 3 given by $[R^2T, R^3SM, ST]$[8] is given by the 8-state graph shown in Fig. 10 (the first label on each edge). The second label shows the image of the group action if this group system is applied to the seed $x_0 = (0,0)^t$ (i.e., the graph is labeled with "group element/image"). Note that x_0 has a nontrivial stabilizer $\langle S, T\rangle$, and that the orbit of x_0 gives the module $\mathbb{Z}_2 \times \mathbb{Z}_4$. The presentation of the orbit system given by this graph is not backward-deterministic, and has states with identical futures. By merging all such states, we obtain a bideterministic minimal graph given by Fig. 11. It then follows that the orbit system in question is a BMP system.It is also easily seen that it is conjugate to the full 4-shift, so that it is a DPDF system as well as CFS. It is not hard to prove that no group system over the above 32-element group alphabet whose minimal graph has just four states generates the subshift in Fig. 11 as its orbit system. However, the smallest group system over the same alphabet that contains the cycle $[R^2T, RS, STM]$ is given by the 8-state graph shown in Fig. 12, by the first set of labels. The orbit system generated by this group system under the action on the same seed x_0 does not collapse. □

Example 10 (An Orbit System Without a BP): Let us now consider the smallest group system over the same alphabet that

[8]This notation is used to represent the bi-infinite sequence

$$\cdots I, I, R^2T, R^3SM, ST, I, I, \cdots.$$

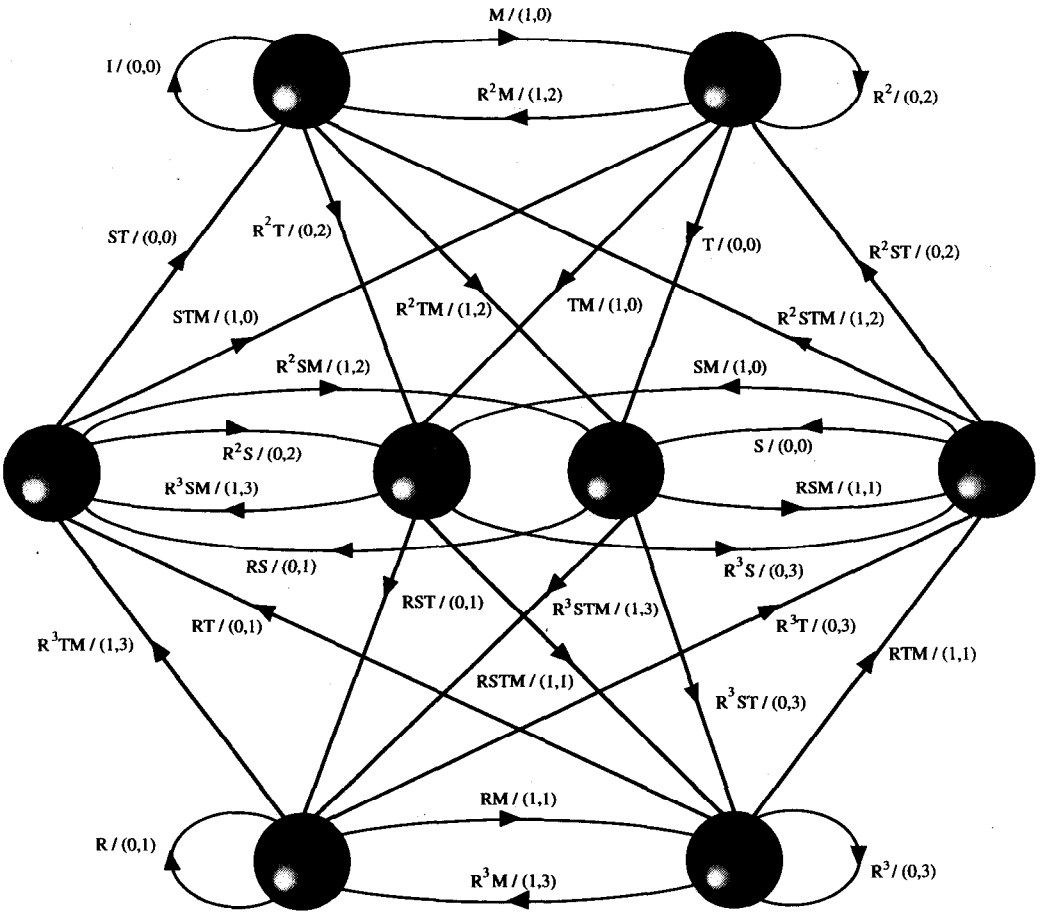

Fig. 10. The group system generated by $[R^2T, R^3SM, ST]$.

contains the finite cycle $[S, RS, T]$. This system is generated by the 8-state graph shown in Fig. 13. Upon application of the same seed $(0,0)^t$ to the group system, we obtain the graph represented by the second set of labels of Fig. 13. This graph is neither backward- nor forward-deterministic, and has certain pairs of states with identical futures and other pairs of states with identical pasts. Hence, the graph may be reduced either by merging states with identical futures or with identical pasts. The former yields the backward-deterministic graph shown in Fig. 14(a) while the latter procedure yields the forward-deterministic graph in Fig. 14(b). These are nonisomorphic as *labeled* graphs, and it follows that the orbit system does not have a bideterministic minimal presentation. Since it may be easily verified that the code is an SFT, it follows that it is not a DPDF system. This may also be checked directly by noting that the images of the left-semi-infinite sequences $\cdots I, I, I$ and $\cdots RM, RM, RM$ share the image of the right-semi-infinite sequence $I, I, I, \cdots$ in their futures, but not the image of the right-semi-infinite sequence $R^3, R^3, R^3 \cdots$. It may, however, be seen the the above orbit system is conjugate to the full 4-shift, so that it is still CFS. □

E. Characterization of Orbit Systems

In the remainder of this section, we will establish the fact that orbit systems are always conjugate to full shifts, thereby providing a complete solution to the problem of their classification. We will first prove that all orbit systems are SFT's by showing that every group code is topologically conjugate to the set of bi-infinite edge sequences of a labeled graph generating it. We will then use a modification of the technique used in the proof of Fact 28 to establish a topological conjugacy between this set of bi-infinite edge sequences and a fullshift, there by demonstrating that every orbit system is CFS.

Let G be a finite group acting on a set X and let $\mathcal{T}(x_0)$ be an orbit system. By passing over to a higher block system if necessary, we may assume that the symbolic dynamic group $\mathcal{T} < G^{\mathbb{Z}}$ is a one-block SFT. Let $(\mathcal{G}, l)$ be the minimal labeled graph generating the subshift $\mathcal{S}$. Let S and E be the state group and the edge set of the directed graph $\mathcal{G}$, and let $l(e) \in G$ denote the label of the edge e, for each $e \in E$. Since the group system is a one-block SFT, it follows that all the edge labels are distinct. The orbit system $\mathcal{T}(x_0)$ is generated by the same graph $\mathcal{G}$ with the new edge labeling l_x, given by

$$l_x(e) = l(e)x_0 \in X$$

for all $e \in E$. The resulting labeled graph $(\mathcal{G}, l_x)$ may not be the minimal presentation of the orbit system $\mathcal{T}(x_0)$, but it has the following useful property:

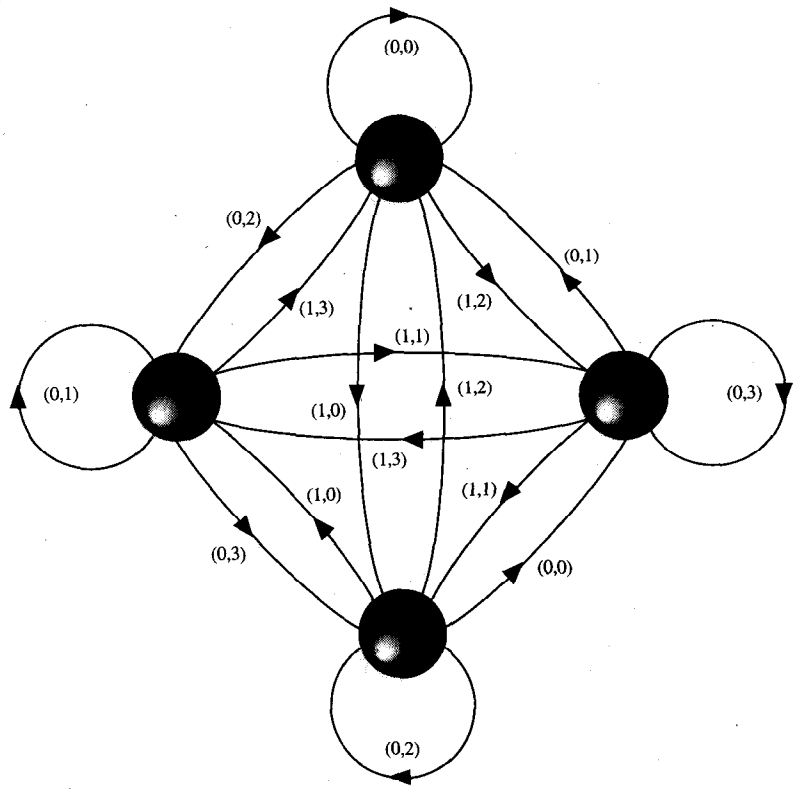

Fig. 11. The minimal BP of the orbit system of Fig. 10.

Fact 31: The futures and pasts of any two states of $(\mathcal{G}, l_x)$ are either identical or disjoint.

See [16]. □

It is possible, however, for two states to have identical futures but disjoint pasts, or the converse. In fact, this is the key difference between group systems and orbit systems.

Let $(\mathcal{G}_r, l_r)$ be the graph obtained from $(\mathcal{G}, l_x)$ by merging all states with identical pasts. The resulting graph still generates the orbit system $\mathcal{T}(x_0)$ and turns out to be very useful in providing insight into its structure. Its utility as a presentation of the orbit system $\mathcal{T}(x_0)$ is further enhanced by the following observation:

Lemma 32: The labeled graph $(\mathcal{G}_r, l_r)$ is deterministic.

Proof of Lemma 32: Suppose the graph $\mathcal{G}_r$ has two edges labeled $a \in X$, with the same initial state s and distinct final states s_1 and s_2. Then, the states s_1 and s_2 share at least one common past. But this is a contradiction, because the construction of $(\mathcal{G}_r, l_r)$ from the graph $(\mathcal{G}, l_x)$, together with Fact 31 ensures that the pasts of any two states of $\mathcal{G}_r$ are mutually disjoint. □

Next, we will show that the orbit system $\mathcal{T}(x_0)$ is conjugate to the set of bi-infinite edge sequences corresponding to paths on the graph G_r, there by establishing the following theorem.

Theorem 33: Every orbit system is an SFT.

Proof of Theorem 33: It is sufficient to show that the 1-block map induced by graph labeling from the edge sequences generated by bi-infinite paths on G_r to the bi-infinite sequences in the orbit system is 1-1. In view of the fact that G_r is deterministic, it is enough to show that every sequence in the orbit system corresponds to a unique bi-infinite state sequence on the graph G_r. Now suppose that two paths with distinct state sequences $\cdots s_{-2}s_{-1}s_0s_1s_2\cdots$ and $\cdots t_{-2}t_{-1}t_0t_1t_2\cdots$ generate the same bi-infinite sequence $x \in \mathcal{T}(x_0)$. Without loss of generality, we may assume $s_0 \neq t_0$. Then we have two distinct states s_0 and t_0 of the graph G_r sharing a common past $x|_{-\infty}^{-1}$, which leads to the same contradiction as before. □

Let ϕ be the labeled graph homomorphism from the graph $(\mathcal{G}, l)$ generating the group system T, on to the graph $(\mathcal{G}_r, l_r)$ generating the orbit system. Let $D < G$ be the stabilizer of the point $x_0 \in X$ with respect to the group action under consideration. Let $S_0^l \subseteq S$ be the set of terminal states of left semi-infinite paths whose edge labels are all drawn from the stabilizer group D. It is easy to see that S_0^l is a subgroup of the state group S, and that two states of the graph $(\mathcal{G}, l_x)$ have the same past iff they lie in the same *left coset* of the subgroup S_0^l. Thus the states of the reduced graph $\mathcal{G}_r$ may be identified with the left cosets of S_0^l, and the graph homomorphism ϕ maps each state s of the graph $\mathcal{G}$ to the left coset sS_0^l. Define H to be the set of labels of all edges in the labeled graph $(\mathcal{G}, l)$ whose initial and final states are in the the subgroup S_0^l. H is a subgroup of the group alphabet G, which we call the *parallel transition group* of the orbit system, and contains a further subgroup $H \cap D$, which we refer to as the *identity subgroup*. Two edges of the graph $(\mathcal{G}, l)$ get mapped to the same edge of $(\mathcal{G}_r, l_r)$ under the graph homomorphism ϕ iff their edge labels lie in the same left coset of the identity subgroup $H \cap D$. Two edges of the graph $(\mathcal{G}, l)$ get mapped to distinct parallel edges on the graph $(\mathcal{G}_r, l_r)$ iff they lie in the same left coset of the parallel transition subgroup H, but in different left cosets of the identity subgroup $H \cap D$. This sheds some light on

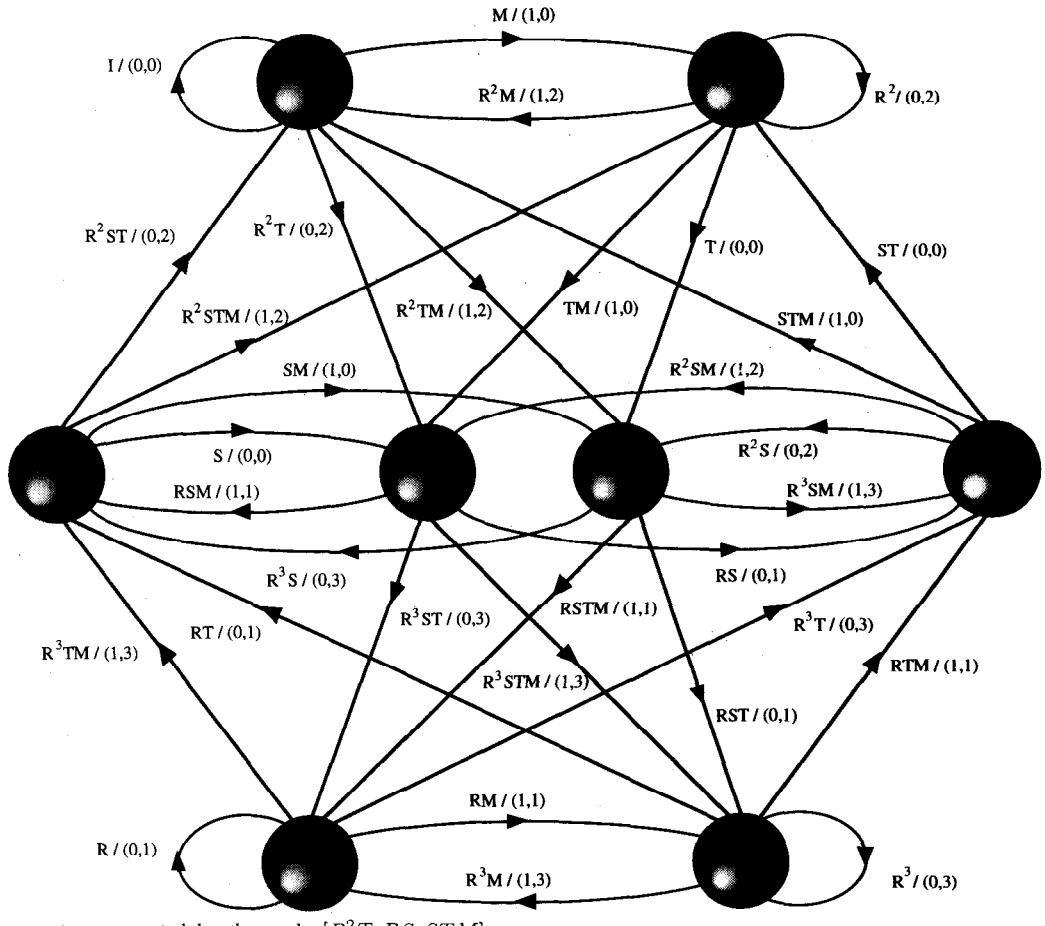

Fig. 12. The group system generated by the cycle $[R^2T, RS, STM]$.

the structure of the deterministic graph $(\mathcal{G}_r, l_r)$ generating the trellis orbit system.

We now come to the most interesting fact about orbit systems.

Theorem 34: Every irreducible orbit system is topologically conjugate to a fullshift.

Proof of Theorem 34: From the preceding facts about the structure of the deterministic graph $(\mathcal{G}_r, l_r)$ generating the orbit system $\mathcal{T}(x_0)$, it follows that whenever there exists an edge from one state to another, there exist exactly $|H|/|H \cap D| \equiv N$ parallel edges between the given pair of states. This shows that the orbit system is conjugate to the product of a full N-shift and the set of all state sequences corresponding to bi-infinite paths on the graph $(\mathcal{G}_r, l_r)$. Let us denote this latter subshift by C'. We will show that C' is actually an orbit system.

Denote by T' the set of all state sequences of the graph $(\mathcal{G}, l)$ corresponding to bi-infinite paths on the graph $\mathcal{G}$. $\mathcal{T}'$ is a group system over the (group) alphabet S. Consider the group action of S on the set Y of all the left cosets of S_0^l induced by left translation, i.e., the group element $s \in S$ acting on the left coset $tS_0^l \in Y$ yields the left coset $(st)S_0^l \in Y$. Let the subgroup $S_0^l \in Y$ be chosen as the seed y_0. It is clear that the subshift C' may be identified with the orbit system $\mathcal{T}'(y_0)$. We have thus shown that the given orbit system is conjugate to a product of a fullshift and another orbit system. Unless the group system $\mathcal{T}$ has a trivial input group, the group system $\mathcal{T}'$ associated with the latter group code has a strictly smaller alphabet. Hence, we may continue the procedure repeatedly, until it is found that the given orbit system is conjugate to a product of several fullshifts and an orbit system $\tilde{C}$ whose underlying group system $\tilde{\mathcal{T}}$ has a trivial input group. But by irreducibility, it follows that the $\tilde{\mathcal{T}}$ and $\tilde{C}$ are full 1-shifts, so that the given orbit system is conjugate to a fullshift. It is a simple matter to see that the size of the fullshift to which the orbit system is conjugate is given by $|F||S_0^l|/|H \cap D|$, where F is the input group of the underlying group system $\mathcal{T}$. □

We summarize our observations in the Venn diagram of Fig. 16; the numbers correspond to the figures in this section. Note that Fig. 15 represents an SFT that is CFS and DPDF but which is not an orbit system. (The latter follows from the nonuniformity of the in- and out-degrees of the graph). This example is included for completeness.

V. Trellis Group Codes for the Gaussian Channel

A trellis group code for the Gaussian channel is obtained when all the ingredients that we have described thus far are brought together. A typical list of constituents consists of

1) a *block isometry group code* $\mathbb{C} \subset \mathbb{R}^n$,
2) a *geometrically uniform partition* $[\mathbb{C}; \mathbb{C}_i]$ of the group code $\mathbb{C}$,

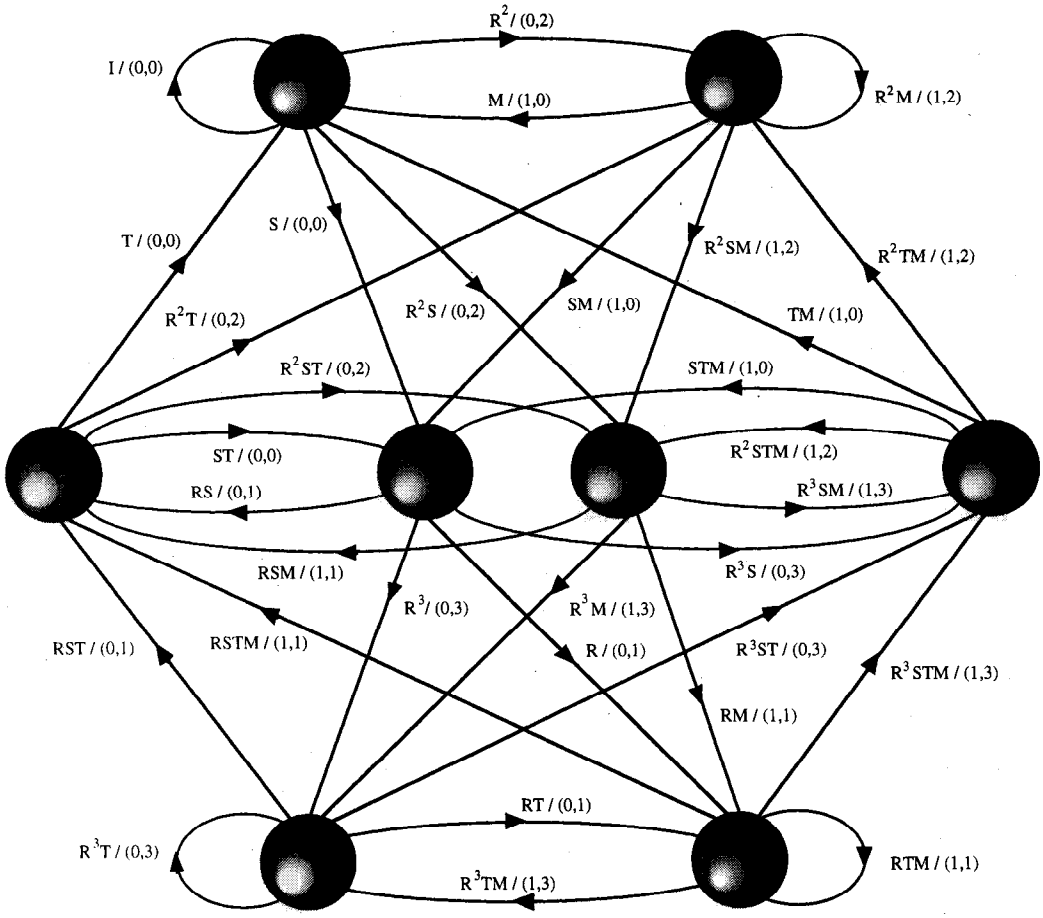

Fig. 13. The group system generated by the cycle $[S, RS, T]$.

3) an *isometric labeling* $l : \mathbb{C} \to \mathcal{L}$ with *label isometry group* $\mathcal{G}_l$,
4) a *symbolic dynamic group* $\Lambda < \mathcal{G}_l^{\mathbb{Z}}$ over the label group $\mathcal{G}_l$, and
5) an *orbit system* $\Lambda(\boldsymbol{x}_0) \subset \mathcal{L}^{\mathbb{Z}}$ for a seed $\boldsymbol{x}_0 \in \mathcal{L}^{\mathbb{Z}}$ (a constant sequence).

The codeword sequences $\boldsymbol{x} \in \mathbb{C}^{\mathbb{Z}}$ are then identified by the necessary and sufficient condition that the label sequence derived from $\boldsymbol{x}$ belong to the orbit system $l(\boldsymbol{x}) \in \Lambda(\boldsymbol{x}_0)$.

In terms of a directed graph (or trellis) description for the code, a minimal graph for the orbit system can be used to describe the set of codewords. For each edge of the orbit graph, substitute the label from $\mathcal{L}$ with the cell of the partition corresponding to the inverse image under the label map l. Notice that it can be assumed that the parallel transition subgroup of the group system is trivial (i.e., each pair of states (s_0, s_1) is connected by at most one directed edge from s_0 to s_1). This follows from the fact that if this were not the case, a coarser partition of the group code $\mathbb{C}$, obtained by the union of the cells associated with the self-loops of the identity state (and the corresponding cosets), would be sufficient to describe the code.

Of course, this class of trellis group code is quite broad and covers a majority of trellis and convolutional codes used in practice. In general, three distinct subclasses exist based on the size of the cells of the partition $|\mathbb{C}_i|$. If the partition is trivial, in the sense that the subcodes consist of only a single point $|\mathbb{C}_i| = 1$, then the distinction between ingredient 1) and 2) is blurred and there is a one-to-one correspondence between the codewords of the code and the orbit system. In terms of the trellis description, every edge of the graph would be labeled by a single point of the group code $\mathbb{C}$. An example of such a system would be QPSK modulation with the standard $\mathbb{Z}_2^2$ labeling and a rate $1/2$ binary convolutional code.

If, on the other hand, the partition is nontrivial yet the cells are finite in cardinality, $1 < |\mathbb{C}_i| < \infty$, then the trellis group code has many (in fact, an infinite number) of codewords for each element of the orbit system. This means, from the encoding point of view, that to encode to a codeword $\boldsymbol{x}$, some data will be associated with the selection of the orbit system sequence, $l(\boldsymbol{x}) \in \Lambda(\boldsymbol{x}_0)$, and the balance of the data is associated with the resolution of the specific point $\boldsymbol{x}$ in the inverse image of the label map $l^{-1}(l(\boldsymbol{x})) \subset \mathbb{C}^{\mathbb{Z}}$. The rate of the encoding is then the sum of the capacity of the orbit system plus the logarithm of the size of the cells of the partition, $\log_2(|\mathbb{C}_i|)$ (bits). In the language of Ungerboeck [7], [8], the encoding maps the "coded bits" to the orbit system, at its capacity, and maps the "uncoded bits" to the specific points in specified cells of the partition. These uncoded bits can be considered the "parallel edge information" since this information is used to resolve the exact path through the trellis once the coded bits determine the state path. An example of

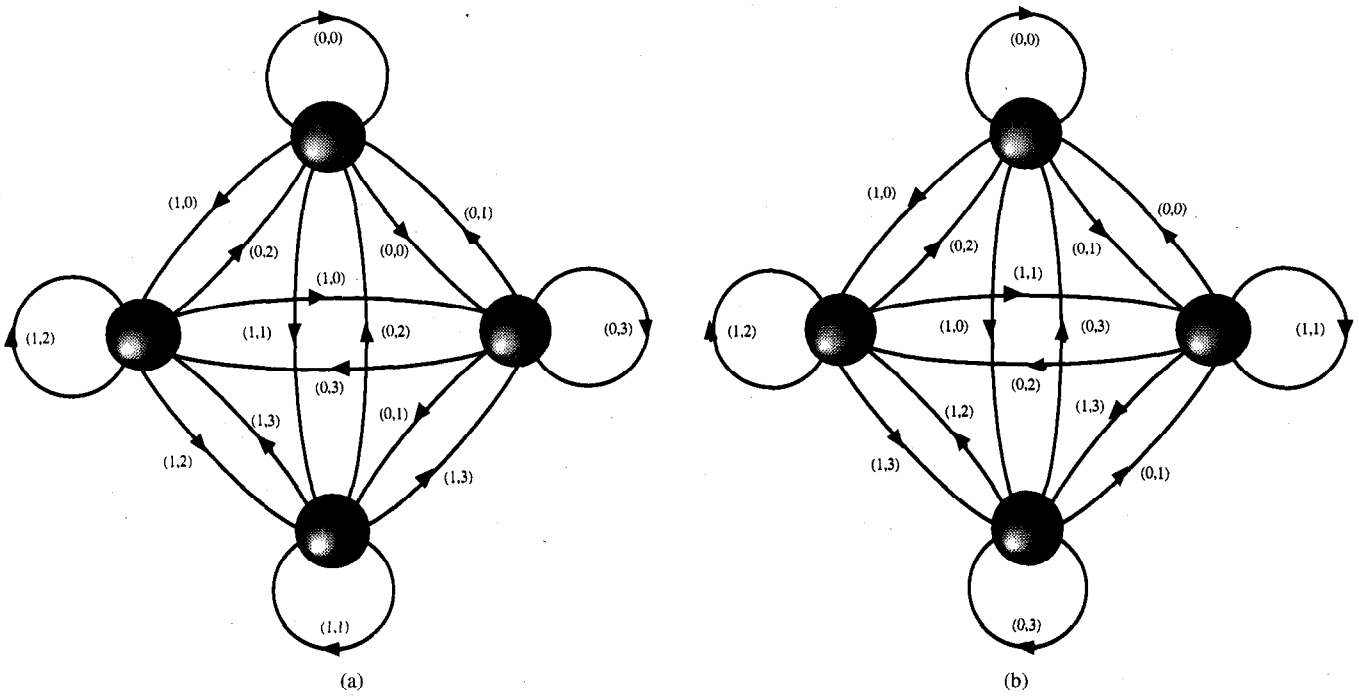

Fig. 14. The backward and forward deterministic minimal graphs of Fig. 13.

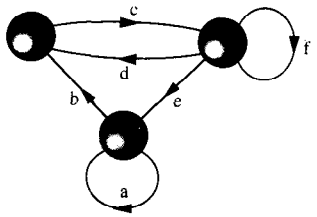

Fig. 15. Example of an SFT that is CFS and DPDF, but not an orbit system.

such a system would be 8-PSK modulation with the standard four-way, antipodal partition with a $\mathbb{Z}_2^2$ labeling of the cells and a rate $1/2$ binary convolutional code.

Finally, when the cells of the partition are infinite in size (e.g., the integer lattice), then the trellis group code has infinite capacity (and infinite power, etc.) and a bounding region $\mathbb{R}$ is used to define a usable subcode. The bounding region $\mathbb{R}$ is defined so that the cardinality of its intersection with each cell is a constant. In that way, $\log_2(|\mathbb{R} \cap \mathbb{C}_i|)$ bits are used as the parallel edge information. In this case, the range of the encoder is not a group code itself, yet inherits many desirable properties of the infinite capacity trellis group code in which it lies. For example, the minimum distance of the code can be lower-bounded by the group code; this is the case, for instance, in trellis-coded QAM modulation. For example, in the v.32bis modem standard, the orbit system is an 8-state system with a capacity of 2 bits and the bounding region is varied so that $\log_2(|\mathbb{R} \cap \mathbb{C}_i|)$ varies from 0 to 4 bits (the rate varies from $(0+2)*2400 = 4800$ to $(4+2)*2400 = 14\,400$ bits/s).

There are two main issues associated with these classes of codes, the *synthesis problem* and the *analysis problem.* In the synthesis problem, one is asked to find attractive codes for a particular application. In this case, the code designer must decide on the five ingredients in order to find an "optimal" tradeoff of parameters, e.g., rate, average and peak power, minimum distance, complexity, etc. Often the system design must satisfy certain coding requirements, such as a rotational invariance constraint, that requires the code be closed under a rotation (e.g., for $\mathbb{R}^2$, require that every codeword, when rotated by 90° component-wise, be a codeword). Such a constraint on a trellis group code is a requirement that the "all-rotation" sequence be a member of the symmetries of the trellis group code. In other words, the code must be realizable as the orbit system of a symbolic dynamic group which includes the all-rotation sequence $\cdots R, R, R, \cdots$.

In the analysis problem, a code is presented to the analyst, in terms of a finite-state machine or similar description, who must try and determine if the code is a trellis group code, and if it is, determine the symmetries of the code. In this case, one must try and determine if there exists a group system Λ that can be used to generate the code. This problem is like solving a puzzle that may or may not have missing pieces.

A. Trellis Group Codes over PSK Signal Sets

In this section, we use the earlier results on group systems over cyclic and dihedral groups in order to characterize trellis group codes over two-dimensional m-PSK signal sets. We will see that any such nontrivial trellis group code must use a four-way partition of the signal set. As a consequence, it follows that there are no nontrivial trellis group codes over the two-dimensional m-PSK signal set that are *strongly rotationally invariant*, i.e., invariant with respect to rotation by an angle 360°/m.

This result implies, for instance, that there are no 90°, rotationally invariant trellis group codes over two-dimensional QPSK. Similarly, there are no 45°, rotationally invariant trellis

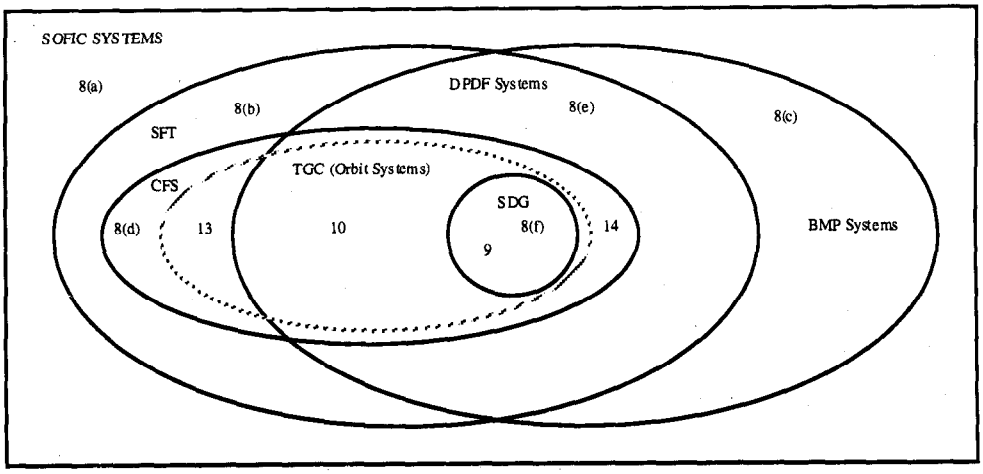

Fig. 16. The relationship between sofic systems; SFT—shift of finite type; BMP—bideterministic minimal presentation; CFS—conjugate to a full shift; DPDF—disjoint past/disjoint future; TGC—trellis group code (orbit system); SDG—symbolic dynamic group.

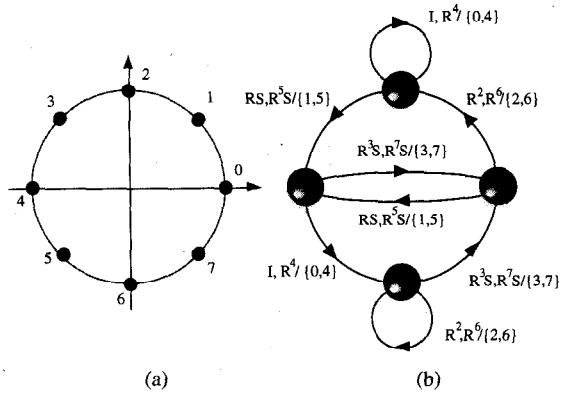

Fig. 17. An example of a 90° RI trellis code over 8-PSK.

group codes over two-dimensional 8-PSK. This, however, does not preclude the existence of codes that satisfy weaker forms of rotational invariance, i.e., invariance with respect to rotations by multiples of 360°/m. The following example from [42] shows a 90°, rotationally invariant trellis group code over two-dimensional 8-PSK. It also follows from our results above that any trellis group code over 8-PSK must be associated with anontrivial four-way partition of the signal set. Since the optimal four-way partition of 8-PSK puts antipodal points in the same cell, it follows that any nontrivial trellis group code over 8-PSK is 180° rotationally invariant. An example of such a code is shown in Fig. 17.

We shall now justify the assertions we made at the beginning of this section. Recall that any irreducible group system $\mathcal{G}$ over a dihedral group may be decomposed into a parallel transition shift $P^{\mathbb{Z}}$, and a quotient shift component $\mathcal{H}$, which is an irreducible group shift over $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. We shall assume that $\mathcal{H}$ is a nontrivial group system over $\mathbb{Z}_2 \oplus \mathbb{Z}_2$; otherwise, the group system $\mathcal{G}$ is a fullshift. We now examine the structure of the orbit system corresponding to the system $\mathcal{G}$ acting on a point $x_0 \in \mathbb{X}$ in the (PSK) signal set. The parallel transition shift $P^{\mathbb{Z}}$ induces a uniform partition $[\mathbb{X}; P(x_0)]$ of the signal set, and the quotient shift may be thought of as acting on the cell $P(x_0)$ of the partition. The quotient group shift is over an abelian group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$, and the orbit system of an abelian group system is itself isomorphic to a group system (since the stabilizer of the group action is normal). Hence, if the stabilizer is trivial, then the orbit system is isomorphic to the group system $\mathcal{G}$. This corresponds to a trellis code over a four-way partition 8-PSK. If the stabilizer is nontrivial, then the orbit system is isomorphic to an irreducible group system over $\mathbb{Z}_2$, which is necessarily a fullshift. Thus any nontrivial code corresponds to a four-way partition of the signal set.

We shall now examine the possibility of strong rotational invariance for such codes. If $\mathbb{D}_m = \langle R_m, S \rangle$ denotes the dihedral group corresponding to the symmetries of m-PSK, then strong rotational invariance implies the existence of a constant sequence $\cdots R_m \cdots$ in some group system generating the given trellis code. From the proof of Theorem 29, it then follows that the parallel transition subgroup contains the cyclic group $\langle R_m^2 \rangle$. But this induces a two-way partition of the signal set, which rules out any nontrivial codes. Hence, we conclude that there are no nontrivial, strongly rotationally invariant trellis group codes over m-PSK constellations in two dimensions.

It follows that in order to realize strong rotational invariance with phase-shift keying, we need to consider higher dimensional m-PSK constellations. We shall now construct a strongly rotationally invariant trellis group code over a four-dimensional QPSK signal set. The signal set consists of 16 points obtained by taking the Cartesian product of two QPSK signal sets, each of whose points are labeled as shown in Fig. 18(a). We denote the symmetries of the 16-point set by ordered pairs of the form $(R^i S^j, R^k S^l)$, with $0 \leq i, k \leq 3$ and $0 \leq j, l \leq 1$, where R represents rotation by 90° while S denotes reflection about the 45° line on the Euclidean plane. We also use I to denote the identity map $(R^0 S^0)$ of the Euclidean plane.

We begin with the subgroup of symmetries of the constellation generated by the elements (R, R), (R^2, I), (S, S), i.e.

$$G = \langle (R, R),\ (R^2, I),\ (S, S) \rangle.$$

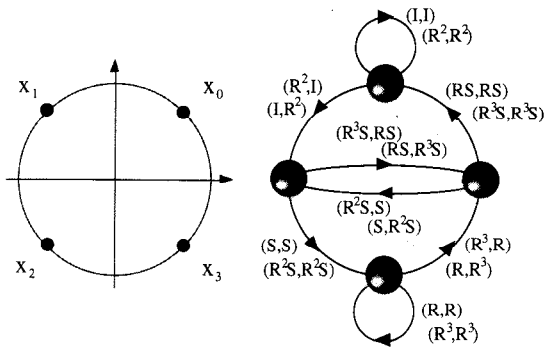

Fig. 18. A QPSK constellation and a four-dimensional group system.

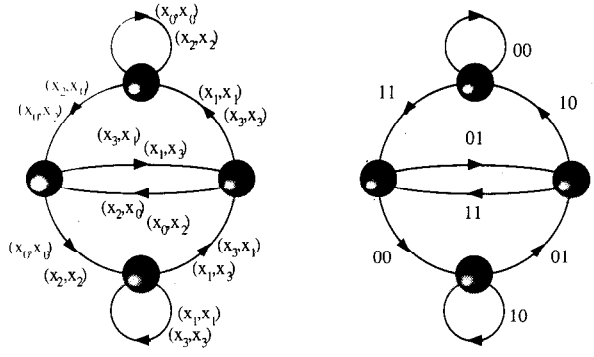

Fig. 19. A 90° rotationally invariant code over four-dimensional QPSK.

Note that G is a nonabelian group of 16 elements. Consider the group system $\mathcal{S}$ over G, generated by the cycle

$$[(R^2, I), (R^2S, S), (RS, RS)].$$

This leads to a graph with 4 states and 2 parallel transitions, depicted in Fig. 18(b). The graph has a self-loop labeled (R, R) on one of its states, which implies that any orbit system generated by $\mathcal{S}$ is strongly rotationally invariant. We apply the group system to the seed $(\boldsymbol{x}_0, \boldsymbol{x}_0)$, generating an orbit system and hence a strongly rotationally invariant trellis group code over four-dimensional QPSK. This code is shown in Fig. 19(a). It is seen that this trellis code has a free distance of 8, which is the maximum possible for any 4-state trellis code over four-dimensional QPSK at the given rate (2 bits/symbol).

There is an alternative way of visualizing this code, in terms of uniform partitions. As mentioned earlier, the parallel transitions in the group system define a group code partition of the signal set, and the quotient group shift acts on the cells of this partition. In this example, the parallel transitions define a GU partition of the 16-point set into eight cells with two points in each cell; some examples of cells of this partition are given by

$$\begin{aligned}\mathbb{C}_0 &= \{(x_0, x_0), (x_2, x_2)\}\\ \mathbb{C}_1 &= \{(x_1, x_3), (x_3, x_1)\}\\ \mathbb{C}_2 &= \{(x_1, x_1), (x_3, x_3)\}\\ \mathbb{C}_3 &= \{(x_0, x_2), (x_2, x_0)\}\end{aligned}$$

etc. The squared Euclidean distance between points in the same cell (intracell distance) is 8, while the minimum squared distance between two distinct cells (intercell distance) is 2.

The quotient group shift corresponding to the group system in Fig. 18(b) is a group system over the abelian group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. As noted earlier, the orbit system of an abelian group system is itself isomorphic to a group system. In this example, the group acting on the cells $\mathbb{C}_i$ of the partition has a nontrivial stabilizer (in fact, a stabilizer isomorphic to $\mathbb{Z}_2$), so that the orbit system is isomorphic to a group system over the alphabet $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. This is equivalent to the convolutional code shown in Fig. 19(b), with generator matrix $G[D] = [1+D+D^2, 1+D]$, where the ordered binary pairs are associated with the cells of the partition in the following manner:

$$00 \leftrightarrow \mathbb{C}_0; \quad 01 \leftrightarrow \mathbb{C}_1 \quad 10 \leftrightarrow \mathbb{C}_2 \quad 11 \leftrightarrow \mathbb{C}_3.$$

Thus the code uses only four of the eight cells in the partition. Since the minimum distance between any two of these four cells is 4, we find that the 4-state code achieves a free distance of 12. Hence, the overall distance of the trellis code is governed by the intracell distance of 8.

Further examples of rotationally invariant trellis group codes over multidimensional m-PSK constellations may be found in [42].

B. Application: Rotationally Invariant Trellis Group Codes for QAM

Rotationally invariant (RI) trellis codes are important whenever the modulation signal set has a rotational symmetry and the transmission system can introduce a phase rotation [38]–[41]. Rotational invariance means that a trellis code is closed under rotation of the individual elements of the signal set for which it labels. As mentioned previously, for a GU trellis code to be RI within the framework of this paper, it is now clear that a necessary and sufficient conditionis that the orbit system is generated from a group system which includes the "all-rotations" sequence. These ideas are illustrated in the next subsections for two-dimwnsional QAM.

1) Group Codes over Upper Triangular, Affine Maps on $\mathbb{Z}_2^n$: In this section, we consider group codes over the set $\mathbb{Z}_2^n$ obtained via a group of upper triangular, affine transformations of the form

$$g(\boldsymbol{z}) = A\boldsymbol{z} + \boldsymbol{b}, \quad \boldsymbol{z} \in \mathbb{Z}_2^n$$

where A is an invertible upper triangular matrix (i.e., 1's on the diagonal) and $\boldsymbol{b} \in \mathbb{Z}_2^n$. The group has $2^{(n^2-n)/2+n}$ elements ($2^{(n^2-n)/2}$ choices for A and 2^n choices for $\boldsymbol{b}$). The group codes of interest are conjugate to the full 2^k shift (i.e., $(\mathbb{Z}_2^k)^{\mathbb{Z}}$) where $k < n$. We discuss the $k = 2$, $n = 3$ case in detail, which has direct applications to the problem of finding interesting rotationally invariant codes in two dimensions, and indicate the framework for general k and n.

Consider an isometric labeling on the label set $\mathbb{Z}_2^3$ with label isometry group

$$\mathcal{G}_l = \{g \mid g(\boldsymbol{z}) = A\boldsymbol{z} + \boldsymbol{b}\}$$

where

$$A = \begin{pmatrix} 1 & a_2 & a_1 \\ 0 & 1 & a_0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{Z}_2^{3\times 3}$$
$$\boldsymbol{b} = (b_2, \quad b_1, \quad b_0)^t \in \mathbb{Z}_2^3, \boldsymbol{z} \in \mathbb{Z}_2^3.$$

This nonabelian group has $|\mathcal{G}_l| = 64$ elements; the elements of the group compose as $g'(g(\boldsymbol{z})) = A\boldsymbol{z} + \boldsymbol{b}$, where

$$A = \begin{pmatrix} 1 & a_2' + a_2 & a_1' + a_1 + a_2' a_0 \\ 0 & 1 & a_0' + a_0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\boldsymbol{b} = \begin{pmatrix} b_2' + b_2 + a_2' b_1 + a_1' b_0 \\ b_1' + b_1 + a_0' b_0 \\ b_0' + b_0 \end{pmatrix}.$$

We are interested in describing group codes over the label set $\mathbb{Z}_2^3$ which encodes $k = 2$ bits, $\mathbb{Z}_2^2$. Let the $(\boldsymbol{x}, \boldsymbol{y}) \in (\mathbb{Z}_2^2)^{\mathbb{Z}}$ be the input sequences and let σ be the shift map. The group system is defined by sliding window maps that generate

$$(a_2, a_1, a_0, b_2, b_1, b_0) \in (\mathbb{Z}_2^6)^{\mathbb{Z}}.$$

The maps are shift-invariant

$$\begin{aligned} a_2(\sigma(\boldsymbol{x}), \sigma(\boldsymbol{y})) &= \sigma(a_2(\boldsymbol{x}, \boldsymbol{y})) \\ a_1(\sigma(\boldsymbol{x}), \sigma(\boldsymbol{y})) &= \sigma(a_1(\boldsymbol{x}, \boldsymbol{y})) \\ &\vdots \\ b_0(\sigma(\boldsymbol{x}), \sigma(\boldsymbol{y})) &= \sigma(b_0(\boldsymbol{x}, \boldsymbol{y})) \end{aligned}$$

and are fixed Boolean functions of a finite number of variables drawn from the $\boldsymbol{x}$ and $\boldsymbol{y}$.

In order to obtain a group code, a seed constant $\boldsymbol{z}_0 \in \mathbb{Z}_2^3$ is selected and the image of the group system acting on $\boldsymbol{z}_0$

$$\begin{aligned} c_3(\boldsymbol{x}, \boldsymbol{y}) &= b_2(\boldsymbol{x}, \boldsymbol{y}) + z_2 + a_2(\boldsymbol{x}, \boldsymbol{y}) z_1 + a_1(\boldsymbol{x}, \boldsymbol{y}) z_0 \\ c_2(\boldsymbol{x}, \boldsymbol{y}) &= b_1(\boldsymbol{x}, \boldsymbol{y}) + z_1 + a_0(\boldsymbol{x}, \boldsymbol{y}) z_0 \\ c_0(\boldsymbol{x}, \boldsymbol{y}) &= b_0(\boldsymbol{x}, \boldsymbol{y}) + z_0 \end{aligned}$$

is the group code. In the important special case where $\boldsymbol{z} = 0$, the code is simply the triple (b_2, b_1, b_0) (this is the case considered subsequently). The group codes of interest in these cases are obtained when the maps from $(\boldsymbol{x}, \boldsymbol{y})$ to (c_2, c_1, c_0) are one-to-one.

The maps from $(\boldsymbol{x}, \boldsymbol{y})$ to $(a_2, a_1, a_0, b_2, b_1, b_0)$ describe a group system iff the system is closed under composition. Closure is obtained when there exists a sliding window map

$$f : (\mathbb{Z}_2^3)^2 \times (\mathbb{Z}_2^3)^2 \to (\mathbb{Z}_2^3)^2, f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}') = (\boldsymbol{x}'', \boldsymbol{y}'')$$

(where $f(\sigma(\boldsymbol{x}), \sigma(\boldsymbol{y}), \sigma(\boldsymbol{x}'), \sigma(\boldsymbol{y}')) = \sigma(f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}'))$) that ensures that the composition equations

$$\begin{aligned} a_2'' &= a_2' + a_2 & b_2'' &= b_2' + b_2 + a_2' b_1 + a_1' b_0 \\ a_1'' &= a_1' + a_1 + a_2' a_0 & b_1'' &= b_1' + b_1 + a_0' b_0 \\ a_0'' &= a_0' + a_0 & b_0'' &= b_0' + b_0 \end{aligned}$$

hold for all $\boldsymbol{x}, \boldsymbol{y} \in (\mathbb{Z}_2^2)^{\mathbb{Z}}$ where $a_2'' = a_2(\boldsymbol{x}, \boldsymbol{y})$, etc. We call this map the *input composition map* f.

2) Rotationally Invariant Trellis Group Codes in Two Dimensions: Consider the isometric labeling for the eight-way Ungerboeck partition in two dimensions based on the label set $\mathbb{Z}_2 \times \mathbb{Z}_4$, as shown in Fig. 7. The label (m, p) represents the binary "magnitude" $m \in \mathbb{Z}_2$ and the four-way "phase" $p \in \mathbb{Z}_4$ of the "windmill" tile. The 32 affine maps given by

$$g\left(\begin{pmatrix} m \\ p \end{pmatrix}\right) = \begin{pmatrix} 1 & a_1 \\ 0 & a_0 \end{pmatrix} \begin{pmatrix} m \\ p \end{pmatrix} + \begin{pmatrix} b_1 \\ b_0 \end{pmatrix}$$

where $a_1, b_1 \in \mathbb{Z}_2$, $b_0 \in \mathbb{Z}_4$ and $a_0 \in \{1, -1\}$, constitute an isometric label group that includes a representative for the 90° rotation of the plane. Rotation corresponds to an increment of the phase component of the label $p \mapsto p + 1$ ($a_1 = b_1 = 0$, $a_0 = b_0 = 1$).

The partition permutation group, $\mathcal{P}\text{pg}$, for the eight-way partition has 64 elements. The entire group is obtained by replacing the windmill labeling with the binary labeling $\boldsymbol{z} = (m, p_1, p_0) \in \mathbb{Z}_2^3$ where $p = 2 * p_1 + p_0$ (i.e., p_1 is the most significant bit (MSB) and p_0 the least significant bit (LSB) of the four-way phase p). Then the isometry label group $\mathcal{G}_l$ represents the entire group $\mathcal{P}\text{pg}$ and takes the form of the affine maps on $\mathbb{Z}_2^3$ as previously discussed. Note that rotation in this label represents the condition $a_2 = a_1 = b_2 = b_1 = 0$, $a_0 = b_0 = 1$; a group code is rotationally invariant if the constant sequences satisfying these conditions is an element of the group system generating the code.

3) A Class of Rotationally Invariant Group Codes: Consider a rate $1/2$ convolutional encoder over $\mathbb{Z}_4$ with polynomial generator

$$\begin{aligned} G(D) &= [g_m(D), g_p(D)] \\ &= [D^k - D^i + 2f_m(D), D^j + 2f_p(D)] \end{aligned}$$

where the binary polynomials

$$f_m(D) = \sum_{l \in L_m} D^l$$

and

$$f_p(D) = \sum_{l \in L_p} D^l.$$

The windmill label is obtained by setting the binary magnitude m to the MSB of the output obtained from the $g_m(D)$ polynomial and the phase as the output from $g_p(D)$. Note that the fact that $(g_p(D))^2 = D^{2j}$ means that the input can be obtained from the phase alone with a sliding window map (i.e., the encoder is "noncatastrophic"). Furthermore, a constant input sequence produces a constant output sequence, with magnitude equal to the the number of terms in the polynomial $f_m(D)$, $|L_m|$, modulo 2; for rotational invariance it is required that this be 0 (i.e., $|L_m|$ is even).

In terms of the binary label, the encoder is expressed by the sliding window equations

$$b_2 = \sigma^k(\boldsymbol{x}) + \sigma^i(\boldsymbol{x}) + \sigma^i(\boldsymbol{y}) + \sum_{l \in L_m} \sigma^l(\boldsymbol{y}) + \sigma^k(\boldsymbol{y})\sigma^i(\boldsymbol{y})$$

$$b_1 = \sigma^j(\boldsymbol{x}) + \sum_{l \in L_p} \sigma^l(\boldsymbol{y})$$

$$b_0 = \sigma^j(\boldsymbol{y}).$$

Note that the equations are linear functions of $(\boldsymbol{x}, \boldsymbol{y})$ with the single exception of the quadratic term $\sigma^k(\boldsymbol{y})\sigma^i(\boldsymbol{y})$, in b_2.

To show that the codes are group codes, we need to exhibit the maps (a_2, a_1, a_0) and the input composition map f. First we take $a_2 = 0$, which would imply that the original $\mathbb{Z}_2 \times \mathbb{Z}_4$ label is sufficient to describe the group code using the 32-element group with the maps

$$\begin{pmatrix} 1 & a_1 \\ 0 & (-1)^{a_0} \end{pmatrix} \begin{pmatrix} m \\ p \end{pmatrix} + \begin{pmatrix} b_2 \\ 2b_1 + b_0 \end{pmatrix}.$$

Take as the composition map f a linear form plus an unknown factor

$$x'' = x + x' + \delta_x \qquad y'' = y + y' + \delta_y$$

and observe that closure

$$\begin{aligned} 0 = \Delta_{b_0} &= b_0 + b_0' - b_0'' \\ &= \sigma^j(y + y' - y - y' - \delta_y) = \sigma^j(\delta_y) \end{aligned}$$

implies $\delta_y = 0$. Now

$$\begin{aligned} 0 = \Delta_{b_1} &= b_1 + b_1^p + a_0' b_0 - b_1'' \\ &= a_0' b_0 - \sigma^j(\delta_x) = a_0' \sigma^j(\boldsymbol{y}) - \sigma^j(\delta_x) \end{aligned}$$

implies $\delta_x = \sigma^{-j}(a_0')\boldsymbol{y}$, which is a quadratic term. Finally, consider the equation

$$\begin{aligned} 0 &= \Delta_{b_2} \\ &= \sigma^k(\boldsymbol{y})\sigma^i(\boldsymbol{y}) + \sigma^k(\boldsymbol{y}')\sigma^i(\boldsymbol{y}') + a_1'\sigma^j(\boldsymbol{y}) \\ &\quad - \sigma^k(\delta_x) - \sigma^i(\delta_x) - \sigma^k(\boldsymbol{y}'')\sigma^i(\boldsymbol{y}'') \\ &= \sigma^k(\boldsymbol{y})\sigma^i(\boldsymbol{y}) + \sigma^k(\boldsymbol{y}')\sigma^i(\boldsymbol{y}') + a_1'\sigma^j(\boldsymbol{y}) - \sigma^k(\sigma^{-j}(a_0')\boldsymbol{y}) \\ &\quad - \sigma^i(\sigma^{-j}(a_0')\boldsymbol{y}) - \sigma^k(\boldsymbol{y} + \boldsymbol{y}')\sigma^i(\boldsymbol{y} + \boldsymbol{y}'). \end{aligned}$$

That is solved when

$$\begin{aligned} a_1'\sigma^j(\boldsymbol{y}) = &\left[\sigma^{k-j}(a_0') + \sigma^i(\boldsymbol{y}')\right]\sigma^k(\boldsymbol{y}) \\ &+ \left[\sigma^{i-j}(a_0') + \sigma^k(\boldsymbol{y}')\right]\sigma^i(\boldsymbol{y}). \end{aligned}$$

This suggests two solutions

$$a_0 = \begin{cases} \sigma^{2i-k}(y), & \text{if } j = i \\ \sigma^{2k-i}(y), & \text{if } j = k \end{cases}$$

$$a_1 = \begin{cases} \sigma^{2i-k}(y) + \sigma^k(y), & \text{if } j = i \\ \sigma^{2k-i}(y) + \sigma^i(y), & \text{if } j = k. \end{cases}$$

For example, if we take

$$G(D) = \left[1 - D, 2 + D + 2D^2\right] \quad (i = 0 \text{ and } j = k = 1)$$

then

$$\begin{aligned} a_2 &= 0 & b_2 &= \boldsymbol{x} + \sigma(\boldsymbol{x}) + \sigma(\boldsymbol{y}) + y\sigma(\boldsymbol{y}) \\ a_1 &= \boldsymbol{y} + \sigma^2(\boldsymbol{y}) & b_1 &= \sigma(\boldsymbol{x}) + \boldsymbol{y} + \sigma^2(\boldsymbol{y}) \\ a_0 &= \sigma^2(\boldsymbol{y}) & b_0 &= \sigma(\boldsymbol{y}) \end{aligned}$$

and we obtain the v.32 code of Fig. 12. When we take

$$\begin{aligned} G(D) &= [1 - D, 2 - D] \\ &= [1 - D, D + 2(1 + D)] \quad (i = 0 \text{ and } j = k = 1) \end{aligned}$$

then

$$\begin{aligned} a_2 &= 0 & b_2 &= \boldsymbol{x} + \sigma(\boldsymbol{x}) + \sigma(\boldsymbol{y}) + y\sigma(\boldsymbol{y}) \\ a_1 &= \boldsymbol{y} + \sigma^2(\boldsymbol{y}) & b_1 &= \sigma(\boldsymbol{x}) + \boldsymbol{y} + \sigma(\boldsymbol{y}) \\ a_0 &= \sigma^2(\boldsymbol{y}) & b_0 &= \sigma(\boldsymbol{y}) \end{aligned}$$

TABLE II
ROTATIONALLY INVARIANT TWO-DIMENSIONAL CODES

$g_m(D)$	$g_p(D)$	Code States	Group States	d^2_{free}	N^2_{free}	Notes
$1 - D$	$2 - D$	4	8	4	4	
$1 - D$	$2 + D + 2D^2$	8	8	5	28	V.32
$1 - D$	$2 - D + 2D^3$	16	16	5	8	
$1 - D$	$2 - D + 2D^3 + 2D^4$	32	32	6	22	
$1 - D^2$	$2 - D^2 + 2D^3 + 2D^4$	64	64	6	2	

and we obtain the code of Fig. 12. Note that these codes only differ in the b_1 term; in the former case, both the group code and group system requires three state (i.e., $\sigma(\boldsymbol{x})$, $\sigma(\boldsymbol{y})$, $\sigma^2(\boldsymbol{y})$) while in the latter, the group code requires only two states (i.e., $\sigma^2(\boldsymbol{y})$ is not required to compute (b_2, b_1, b_0)) as shown in Fig. 11.

Table II describes examples of the best codes of this class.

VI. CONCLUSION AND OPEN QUESTIONS

In this paper we have presented a unified view of geometrically uniform trellis codes as an extension of Slepian's group codes. Trellis group codes, and in fact any code described as the orbit of some group in sequence space, are naturally described using the language and techniques of symbolic dynamics.

The theory of trellis group codes furnishes several interesting analysis- and synthesis-related problems. Problems of an analytic nature involve finding efficient algorithms to determine whether a given trellis code is an orbit system or not. Some of the results in this paper on the symbolic dynamical characterization of trellis group codes (i.e., orbit systems) may be strengthened further, thereby leading to an inductive procedure that performs the desired computation. More generally, we would like algorithms for the construction of the symbolic dynamic group that represents all the symmetries of a given trellis code. This is of particular interest in the design of rotationally invariant codes that may or may not be trellis group codes.

There are still many open questions relating to the design of trellis group codes for the Gaussian channel, particularly relating to their distance properties. A partial answer to this problem involves enumeration of group systems up to the desired complexity by means of the cycles that generate them. One of the drawbacks of this approach is that the minimal graph of the group system may be much larger than that of an orbit system it generates. This leads to the question of whether trellis group codes with the best distance properties (for a given number of encoder states) can come from group systems that "collapse," i.e., does the presence of stabilizers in the generators of a group system "cost" distance.

APPENDIX

A. Semi-Direct Products

A group $\mathcal{C} = \mathcal{B} \rtimes \mathcal{A}$ is a *semidirect product* if

i) $\mathcal{B}$ is a normal subgroup of $\mathcal{C}$ $(\mathcal{B} \lhd \mathcal{C})$.

ii) $\mathcal{A}$ is a subgroup of $\mathcal{C}$ $(\mathcal{A} < \mathcal{C})$.

iii) Every element $c \in \mathcal{C}$ is can be uniquely written $c = a*b$, $b \in \mathcal{B}$, $a \in \mathcal{A}$ (i.e., $\mathcal{B} \cap \mathcal{A} = \{I\}$, $\mathcal{A} * \mathcal{B} = \mathcal{C}$).

In this case, if $c_1, c_2 \in \mathcal{C}$, then $c_1 = a_1 * b_1$, $c_2 = a_2 * b_2$ (where $*$ is the group operation), and

$$\begin{aligned} c_3 = c_1 * c_2 &= a_1 * b_1 * a_2 * b_2 \\ &= (a_1 * a_2) * (a_2^{-1} * b_1 * a_2 * b_2) = a_3 * b_3 \end{aligned}$$

where $a_3 = a_1 * a_2 \in \mathcal{A}$ and $b_3 = a_2^{-1} * b_1 * a_2 * b_2 \in \mathcal{B}$ (since $\mathcal{B}$ is Normal). Note that if it is always the case that $a_2^{-1} * b_1 * a_2 = b_1$ (e.g., as is always the case in an Abelian group) then $\mathcal{C} = \mathcal{B} \times \mathcal{A}$ is a *direct product.*

For example, if

$$\mathcal{C} = \{f \mid f(\boldsymbol{x}) = (A\boldsymbol{x}) + \boldsymbol{b}\}$$

is a group of affine transformations, then the translation subgroup

$$\mathcal{B} = \{f \mid f(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}\}$$

is normal in $\mathcal{C}$, the linear subgroup

$$\mathcal{A} = \{f \mid f(\boldsymbol{x}) = A\boldsymbol{x}\}$$

has only the identity map in common with $\mathcal{B}$, and the group $\mathcal{C} = \mathcal{B} \rtimes \mathcal{A}$ since

$$\begin{aligned} f_2(f_1(\boldsymbol{x})) &= (A_2(A_1\boldsymbol{x} + \boldsymbol{b}_1)) + \boldsymbol{b}_2 \\ &= (A_2A_1\boldsymbol{x}) + (A_2\boldsymbol{b}_1 + \boldsymbol{b}_2) \\ &= (A_3\boldsymbol{x}) + \boldsymbol{b}_3. \end{aligned}$$

B. Proofs

Proof of Proposition 24: In view of Fact 22, we may suppose without loss of generality that the bideterministic presentation is irreducible. Let $\mathcal{G}$ be any such presentation of an irreducible sofic system $\mathcal{S}$, and let $\mathcal{G}_0$ be the labeled graph obtained by merging all states of $\mathcal{G}$ with identical futures. Then from Fact 23, $\mathcal{G}_0$ is the right Fischer cover of $\mathcal{S}$ and is the image of $\mathcal{G}$ under a label preserving graph homomorphism ϕ. Suppose that $\mathcal{G}_0$ is not bideterministic (or else we are done) and that a state s of $\mathcal{G}_0$ has two incoming edges e and f with the same label a. Since $\mathcal{G}_0$ is irreducible, there are paths p and q originating from state s such that the terminal states of p and q coincide with the initial states of e and f, respectively. Let $\phi^{-1}(s)$ be the subset of states of the graph $\mathcal{G}$ which get mapped to the state s under the graph homomorphism, and let n be its cardinality. Since all the states in $\phi^{-1}(s)$ have the same future, each one of them has pre-images of the paths p and q originating at them. From the minimality of $\mathcal{G}_0$ and bideterministic nature of $\mathcal{G}$, it follows that all these paths terminate at distinct states of the graph $\mathcal{G}$. This means that there are at least $2n$ distinct edges in $\mathcal{G}$ with the same label a, each of which terminates in one of the n states in $\phi^{-1}(s)$. This contradicts the assumption that $\mathcal{G}$ is bideterministic, thereby establishing the desired result. □

Proof of Proposition 25: ($\Rightarrow$) Let $\mathcal{G}$ be the right Fischer cover generating an irreducible sofic system $\mathcal{S}$ with the DPDF property. Since the future of every state of $\mathcal{G}$ coincides with the future of some left semi-infinite sequence of the DPDF system $\mathcal{S}$, and since by minimality, no two states of $\mathcal{G}$ have the same future, it follows that the states of $\mathcal{G}$ have *disjoint* futures. This implies that the graph $\mathcal{G}$ is backward-deterministic because, if a state has two incoming edges with the same label, then the initial states of the two edges have a common future, which contradicts the minimality of $\mathcal{G}$. This shows that $\mathcal{S}$ is a BMP system.

Now suppose that $\mathcal{S}$ is not an SFT. Then there are two bi-infinite paths on the graph $\mathcal{G}$ which generate the same label sequence. Hence we may choose two right-semi-infinite paths with distinct starting states s and t, producing the same right-semi-infinite label sequence x. Then, the right-semi-infinite sequence x lies in the future of the states s as well as t. But this is impossible unless $s = t$ because the states of $\mathcal{G}$ disjoint futures. Thus contradiction leads to the conclusion that $\mathcal{S}$ is an SFT.

($\Leftarrow$) Let $\mathcal{G}$ be the bideterministic Fischer cover of the sofic system $\mathcal{S}$. In view of the fact that the futures of the states of $\mathcal{G}$ correspond to futures of left-semi-infinite sequences of $\mathcal{S}$, it suffices to show that the states of $\mathcal{G}$ have disjoint futures. Suppose there are right-semi-infinite paths originating from two distinct states of the graph generating the same label sequence. Since the graph is backward-deterministic, the two paths never pass through the same state at the same time. But this means that there are two distinct cycles on $\mathcal{G}$ generating the same string, which is impossible since $\mathcal{S}$ is an SFT. □

Proof of Theorem 29: Let $\mathcal{S}$ be an irreducible group system over the alphabet $G = \mathbb{Z}_n$. Then $\mathcal{S}$ is generated by the set of all its cycles.

Suppose that the assertion of the theorem is true for all cyclic groups $\mathbb{Z}_m$ with $m < n$. If $n = p$ is a prime number, then we have $\mathcal{S} = \mathbb{Z}_n^{\mathbb{Z}}$, since any nontrivial cycle of $\mathcal{S}$ may be identified with a nonzero polynomial over the prime field $\mathbb{F}_p$, and any such polynomial may be inverted in the the field of formal Laurent series over $\mathbb{F}_p$.

Now suppose that n is a composite number, and let $1 < k < n$ be any integer that divides n. Define $\mathcal{T}$ to be the irreducible group system obtained by raising every sequence of $\mathcal{S}$ to its kth power. Then $\mathcal{T}$ is isomorphic to an irreducible group system over the cyclic group $\mathbb{Z}_m$ with $m < n$, and is hence a full shift isomorphic to $\mathbb{Z}_{m'}^{\mathbb{Z}}$, for some integer m' that divides m. We may assume that $m' > 1$ because otherwise it follows that $\mathcal{S}$ is isomorphic to an irreducible group system over the smaller alphabet $\mathbb{Z}_{\frac{n}{k}}$. Now, $\mathcal{T}$ is a normal subgroup of $\mathcal{S}$, and the quotient group $\mathcal{Q} \equiv \mathcal{S}/\mathcal{T}$ is isomorphic to an irreducible subshift over a cyclic group $\mathbb{Z}_r$ with $n = rm'$. Hence, $\mathcal{Q}$ is also a fullshift. Finally, the group system $\mathcal{S}$ is isomorphict o the direct product of the normal subgroup $\mathcal{T}$ and the quotient group $\mathcal{Q}$. The result now follows by induction. □

Proof of Theorem 30: Let $\mathcal{S}$ be an irreducible group system over the dihedral group $\mathbb{D}_m \cong \langle R_m, S\rangle$. Let $\mathcal{T}$ be the group system generated by the cycles obtained by *squaring* each cycle in $\mathcal{S}$. Then $\mathcal{T}$ is an irreducible group system over the cyclic group $\langle R_m\rangle$ and hence is isomorphic to a fullshift by

Theorem 29. Let $\mathcal{T}$ be the fullshift over $\langle R_m^r \rangle$ for some positive integer r that divides m. Since the cyclic group $\langle R_m^r \rangle$ is normal in the dihedral group $\mathbb{D}_m$, it follows that the fullshift $\mathcal{T}$ is normal in the group system $\mathcal{S}$. The quotient group $\mathcal{Q} = \mathcal{S}/\mathcal{T}$ is isomorphic to an irreducible group system over the dihedral group $\mathbb{D}_r (\cong \mathbb{D}_m/\langle R_m^r \rangle)$. Also, from the definition of $\mathcal{T}$, it follows that every cycle of $\mathcal{Q}$ is of order 2, so that $\mathcal{Q}$ is actually a group system over a subgroup $G' < \mathbb{D}_r$ consisting only of elements of order 2. But any such subgroup of the dihedral group $\mathbb{D}_r$ is isomorphic to $\mathbb{Z}_2$ or $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. Note that if $G' \cong \mathbb{Z}_2$, then the original group system $\mathcal{S}$ is itself a fullshift over an abelian subgroup ($\mathbb{Z}_2$ or $\mathbb{Z}_2 \oplus \mathbb{Z}_2$) of the dihedral group.

Note that $\mathcal{Q}$ in the proof above is not necessarily the quotient group shift $\mathcal{S}$, since the fullshift $\mathcal{T}$ could be a proper subset of the parallel transition shift $\mathcal{P}^{\mathbb{Z}}$. □

ACKNOWLEDGMENT

The authors gratefully acknowledge input from several colleagues, in particular: G. D. Forney, A. Loeliger, B. Marcus, T. Mittelholzer, and M. Trott. They also wish to thank the anonymous reviewers whose comments helped improve the original manuscript.

REFERENCES

[1] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, Apr. 1968.
[2] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.
[3] E. Biglieri and M. Elia, "Multidimensional modulation and coding for bandlimited digital channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 803–809, July 1988.
[4] 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, Mar. 1987.
[5] G. D. Forney, Jr., "Coset codes-part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.
[6] ——, "Coset codes-part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1151–1187, Sept. 1988.
[7] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
[8] ——, "Trellis-coded modulation with redundant signal sets—Part I: Introduction—Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, pp. 5–21, Feb. 1987.
[9] H. Wielandt, *Finite Permutation Groups*. New York: Academic Press, 1964.
[10] R. L. Adler and B. Marcus, "Topological entropy and equivalence of dynamical systems," *Amer. Math. Soc.*, vol. 20, no. 219, 1979.
[11] R. L. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes: An application of symbolic dynamics to information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 5–22, Jan. 1983.
[12] B. Marcus, "Sofic systems and encoding data," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 366–377, May 1985.
[13] B. Kitchens, "Expansive dynamics on zero-dimensional groups," *Ergodic Theory Dyn. Syst.*, vol. 7, pp. 249–261, 1987.
[14] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Sel. Areas Commun.*, vol. 10, pp. 5–37, Jan. 1992.
[15] J. C. Willems, "Models of dynamics," in *Dynamics Reported*, vol. 2. New York: Wiley, 1989, pp. 171–266.
[16] M. D. Trott, "The algebraic structure of trellis codes," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1992.
[17] H.-A. Loeliger and T. Mittelholzer, "Convolutional codes over groups," submitted to *IEEE Trans. Inform. Theory.*
[18] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: state spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.
[19] J. J. Rotman, *An Introduction to the Theory of Groups*. Dubuque, IA: Wm. C. Brown, 1988.
[20] J. M. Hall, *The Theory of Groups*. New York: Chelsea, 1976.
[21] I. N. Herstein, *Abstract Algebra*. London, UK: Macmillan, 1986.
[22] I. Ingemarsson, "Group codes for the Gaussian channel," in *Topics in Coding Theory*, no. 128 in *Lecture Notes in Control and Information Sciences*. Berlin, Germany: Springer-Verlag, 1989, pp. 73–108.
[23] W. Ledermann and E. S. Vajda, *Handbook of Applicable Mathematics*, vol. V: *Combinatorics and Geometry*, pt B. New York: John, 1985.
[24] G. D. Forney, Jr., "Progress in geometrically uniform codes," in *Proc. 6th Joint Swedish-USSR Inr. Workshop on Information Theory* (Mölle, Sweden), 1993, pp. 16–20.
[25] R. Buzz, "Uniformity of nonlinear trellis codes," in *Proc.5th Tirrentia Workshop on Digital Communications* (Pisa, Italy), Sept. 1991.
[26] H.-A. Loelinger, "Signal sets matched to groups," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1675–1682, Nov. 1991.
[27] N. T. Sindhushayana, "Symbolic dynamics, automata theory and the theory of coding: A comprehensive study and applications," M.S. thesis, Cornell Univ., Ithaca, NY, 1992.
[28] N. T. Sindhushayana and C. Heegard, "Symbolic dynamics and automata theory, with applications to constraint coding," submitted to *IEEE Trans. Inform. Theory*
[29] T. Mittelholzer, H.-A. Loeliger, G. D. Forney, Jr., and M. D. Trott, "Minimality and observability of group systems," preprint.
[30] G. A. Hedlund, "Endomorphisms and automorphisms of the shift dynamical system," *Math. Syst. Theory*, no. 3, pp. 320–375, 1969.
[31] A. Khayrallah and D. L. Neuhoff, "Subshift models and finite-state codes for input-constrained noiseless channels: A tutorial," Udel-EE Tech. Rep. 90–9–1, Univ. Mich., Ann Arbor, Sept. 1990.
[32] M. Boyle, B. Kitchens, and B. Marcus, "A note on minimal covers for sofic systems," *Amer. Math. Soc.*, vol. 95, pp. 403–411, Nov. 1985.
[33] R. Fischer, "Graphs and symbolic dynamics," in *Topics in Information Theory*, vol. 16 of *Colloquia Mathematica Societatis János Bolyai*. Keszthely: Hungary, 1975, pp. 229–244.
[34] N. Jonoska and B. Marcus, "Minimal presentations for irreducible sofic shifts," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1818–1827, Nov. 1994.
[35] B. H. Marcus and R. M. Roth, "Bounds on the number of states in encoder graphs for input-constrained channels," *IEEE Trans. Inform. Theory*, vol. 37, pp. 742–758, May 1991.
[36] N. T. Sindhushayana and C. Heegard, "Symbolic dynamic groups and generators," presented at the IEEE-IT Workshop, Mt. Fuji, Japan, June 1993.
[37] E. J. Rossin and C. Heegard, "Rotationally invariant trellis codes with a linear structure," in *Proc. 26th Conf. on Information Sciences and Systems* (Princeton Univ., Princeton, NJ), 1992.
[38] L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space—Parts I and II," *IEEE J. Sel. Areas Commun.*, vol. SAC-2, pp. 659–686, Sept. 1984.
[39] ——, "Rotationally invariant trellis-coded modulations with multidimensional m-psk," *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 1281–1295, Dec. 1989.
[40] S. S. Pietrobon, R. H. Deng, A. Lafanechère, G. Ungerboeck, and D. J. Costello, Jr., "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 63–89, Jan. 1990.
[41] S. S. Pietrobon, G. Ungerboeck, L. C. Perez, and D. J. Costello, Jr., "Rotationally invariant, nonlinear trellis codes for two-dimensional modulation," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1773–1791, Nov. 1994.
[42] S. Benedetto, R. Garello, M. Mondin, and G. Montorsi, "Geometrically uniform partitions of $L\times$ MPSK constellations and related binary trellis codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1773–1798, Nov. 1993.