Group-theoretic construction and analysis of generalized coset codes for symmetric/asymmetric distributed source coding.

S. Sandeep Pradhan and
Kannan Ramchandran[1]
Department of Electrical
Engineering and Computer Science
University California, Berkeley
e-mail:
{pradhan5,kannanr}@eecs.berkeley.edu

Abstract — **In this paper we consider the problem of distributed source coding with symmetric and asymmetric rates of transmission with a constructive and practical approach, referred to as DIstributed Source Coding Using Syndromes (DISCUS). We employ generalized coset codes constructed in a group-theoretic setting, for this approach and show that the performance of symmetric encoding systems matches that of asymmetric systems. This approach is very general and encompasses large class of coding techniques such as trellis codes and lattice codes. We give several examples for the illustration of the concepts.**

I. Introduction

Remote encoding of correlated distributed sources (referred to as distributed source coding) was first considered in [1] where it was shown primarily that any pair of discrete memoryless sources X and Y can be separately (remotely) compressed at a rate approaching their joint entropy [2]. This was extended to the case of encoding of sources with a fidelity criterion in [3, 4], where source encoding in the presence of side information (in the form of a correlated source) at the decoder is considered: this was referred to as source coding with side information. Another extension of distributed source coding in a sensor network setting was tackled in [5], where the encoders observe independent noisy versions of a source X and separately encode their observed data to a remote joint decoder to get the best joint estimate of X. Further extensions are studied in [6].

Practical construction of encoders and decoders for such problems was considered in [7], where a trellis-based code construction was formulated and analyzed under the name of DIstributed Source Coding Using Syndromes (DISCUS). The construction in [7] is an asymmetric version of the problem where one of the correlated sources is available at the decoder with "full resolution" and the other source is encoded with finite rate. This problem has been tackled in the more general symmetric setting in [8], where a constructive approach to distributed encoding of correlated sources with arbitrary achievable rate pairs, directly based on linear codes, was introduced. In [8], a trellis- and lattice-based construction of codes for symmetric encoding of correlated sources with a fidelity criterion was introduced and their performance was analyzed.

In this paper we formulate the distributed source coding problem in a general setting, which encompasses the trellis and lattice codes, referred to as coset codes. Coset codes were proposed in [9] where it was shown that trellis codes and lattice codes are special cases of coset codes. Further, their geometric properties were analyzed in the more general setting of generalized coset codes in [10]. It was shown that most classes of signal space codes are generalized coset codes and geometrically uniform [10].

In this paper we construct and analyze generalized coset codes for distributed source coding in both asymmetric and symmetric settings. A key result that will be proved in this paper is that *the performance of symmetric encoding matches that of asymmetric encoding with the same computational complexity.* It should be noted that the encoder of each source sends only partial information, which when combined with similar partial information from the other correlated sources at the decoder, can be used to optimally reconstruct the source under consideration.

This paper is organized as follows. Section 2 gives the preliminaries and illustrative examples. Section 3 gives the construction and analysis of the system based on generalized coset codes. Section 4 deals with some special cases. Section 5 provides the conclusions.

II. Preliminaries

A Preliminaries of coset codes:

We follow the group-theoretic [11] formulation of generalized coset codes as given in [10] and assume familiarity with this. We briefly give the necessary definitions for completeness. We follow the definitions of isometry and symmetry groups as given in [10]. A geometric figure S is said to be geometrically uniform if for any two points $x, y \in S$, there exists an isometry u_{xy} such that $u_{xy}(x) = y, u_{xy}(S) = S$. A generating group $U(S)$ of S is a sub-group of the symmetry group that is minimally sufficient to generate S from an arbitrary point $s_0 \in S$. Thus S can be represented as

$$S = \bigcup_{u \in U(S)} u(s_0), \tag{1}$$

for some initial point $s_0 \in S$. A geometrically uniform partition S/S' is a partition of a geometrically uniform signal set S with a generating group $U(S)$ that is induced by a normal subgroup U' (also referred to as $U(S')$) of $U(S)$. A label group A for a geometrically uniform partition S/S' is a group that is isomorphic to the quotient group $U(S)/U(S')$. Thus the elements of A can be used to label the elements of the quotient group $U(S)/U(S')$ which is determined by the map $m : A \to S/S'$. We consider only Abelian [10, 11] label groups.

[1] This work was supported in part by DARPA grant F29601-99-1-0169.

The subsets of S in this partition are associated with an element of A. Thus we have

$$S'(c) = \bigcup_{u \in U(S')} u_c[u(s_0)], \tag{2}$$

for any element of $c \in A$, and $u_c \in U(S)$ is the coset-representative of $S'(c)$. Consider the following example for illustration.

Example 1: Let $S = \mathbb{Z} + 1/2$ as shown in Fig. 1 and let the generating group be $U(S) = (V.T(2\mathbb{Z}))$, where $T(2\mathbb{Z})$ is the group of translation by multiples of 2 [10] and V is a group with elements in the isometry set $\{I, R_-\}$, where I is the identity function, and R_- is the reflection about the origin with composition as the group operation. Let the normal subgroup $U(S')$ of $U(S)$ be given by: $U(S') = (T(4\mathbb{Z}))$. Clearly S/S' is isomorphic to $\mathbb{Z}_2^2$. Thus when $s_0 = 0.5$, S' is as shown in Fig. 1.

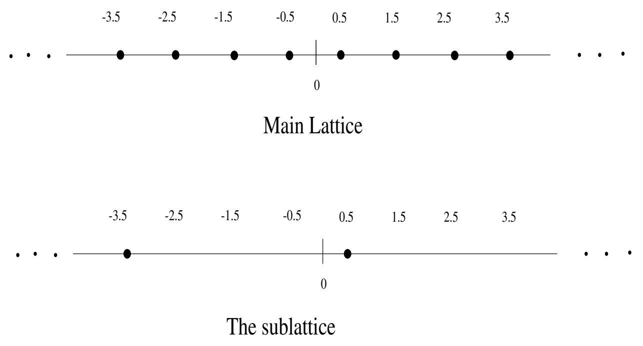

Figure 1: An example where $S = \mathbb{Z} + 1/2$ with the generating group given by $(V.T(2\mathbb{Z}))$, where $T(2\mathbb{Z})$ is the group of integer translation and V is a group with elements in the isometry set $\{I, R_-\}$ with I as identity and R_- as reflection about the origin, the subgroup $U(S')$ of $U(S)$ is given by: $U(S') = T(4\mathbb{Z})$. The sublattice S' with the initial point $s_0 = 0.5$ is shown.

A generalized coset code $\mathbb{C}(S/S'; \mathcal{C})$ is the set of all sequences $\{\mathbf{s} \in \mathbf{m}(\mathbf{c}), \mathbf{c} \in \mathcal{C}\}$, where $\mathcal{C}$ is a subgroup of A^I, A is the label group of S/S', I is the index set, $\mathbf{m} : A^I \to (S/S')^I$ is the sequence extension of the map $m : A \to S/S'$. This is illustrated in Fig. 2, where an encoder for such a code sequence is shown. The sequences $\mathbf{c} \in \mathcal{C}$ select the sequences of cosets S/S', and a part of the indexing bits select the elements in the chosen coset.

Thus the generalized coset code can be represented as

$$\mathbb{C}(S/S'; \mathcal{C}) = \bigcup_{\mathrm{c} \in \mathcal{C}} \mathbf{m}(\mathbf{c}) = \bigcup_{\mathrm{c} \in \mathcal{C}} u_{\mathrm{c}}\left((S')^I\right)$$

$$= \bigcup_{\mathrm{c} \in \mathcal{C}} \bigotimes_{i \in I} \left(\bigcup_{u \in U(S')} u_{c_i}(u(s_0)) \right), \tag{3}$$

where $\mathbf{c} = \bigotimes_{i \in I} c_i$. Any coset of this code can be represented as

$$\mathbb{C}(S/S'; \mathcal{C} + \mathbf{a}) = \bigcup_{\mathrm{c} \in \mathcal{C}} u_{\mathrm{a}}\left[u_{\mathrm{c}}\left((S')^I\right)\right]$$

$$= \bigcup_{\mathrm{c} \in \mathcal{C}} \bigotimes_{i \in I} \left(\bigcup_{u \in U(S')} u_{a_i}[u_{c_i}(u(s_0))] \right). \tag{4}$$

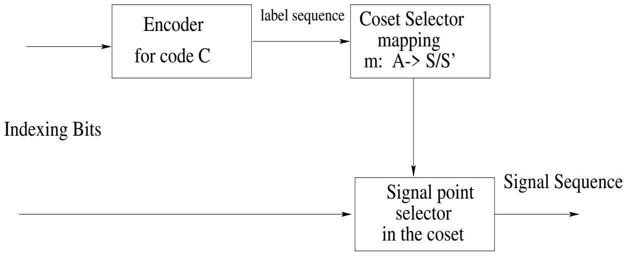

Figure 2: Illustration of a generating mechanism for generalized coset codes: $\mathcal{C}$ selects the sequence of cosets S/S' and some part of the indexing bits select the points in the given coset.

Thus it can be seen that trellis and lattice codes are the special case of generalized coset codes.

B Preliminaries of distributed source coding for discrete sources:

We consider an example for illustration of the basic concepts. Consider[1] a pair of correlated discrete memoryless sources X and Y such that $X, Y \in \{0,1\}^n$ and $d_H(X,Y) \leq t$ where $d_H(.,.)$ is the Hamming distance and let us also suppose that $\sum_{i=0}^{t} \binom{n}{i} = 2^k$ for some integer k (sphere packing bound [12]). According to [1], X and Y can be separately compressed at rate pairs R_1, R_2 given by

$$R_1, R_2 \geq n - k, R_1 + R_2 \geq 2n - k. \tag{5}$$

This region is shown in Fig. 3.

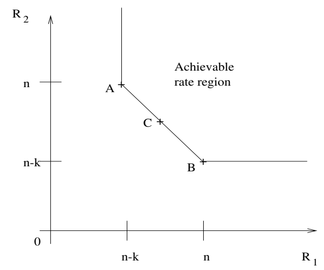

Figure 3: Achievable rate region of two correlated sources: any rate pair (R_1, R_2) can be achieved if $R_1, R_2 \geq n - k$ and $R_1 + R_2 \geq 2n - k$. The sources considered here are $X, Y \in \{0,1\}^n$ and $d_H(X,Y) \leq t$.

Let us consider a system based on $(n, k, 2t+1)$ linear binary code, $\mathcal{C}$, to achieve these points on the rate region. The point $(R_1, R_2) = (n-k, n)$ can be easily achieved [13, 7] using a system where Y sends its data using n bits and X sends the syndrome $\mathbf{s} = \mathbf{Hx}$ where $\mathbf{x}$ is the observed outcome of X, $\mathbf{H}$

[1] Without loss of generality, we use two sources to illustrate the concepts and generalization to more than two sources is straightforward.

is the parity check matrix of $\mathcal{C}$.
Theorem: All the integral rate pairs on (5) can be achieved using $\mathcal{C}$.
Proof: Let $\mathbf{G}$ be the generator matrix of $\mathcal{C}$. Let $\mathbf{G}1$ and $\mathbf{G}2$ be $n - R_1 \times n$ and $n - R_2 \times n$ matrices respectively, obtained by a partition of k rows of $\mathbf{G}$. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be the linear codes associated with $\mathbf{G}1$ and $\mathbf{G}2$ respectively. The encoders associated with X and Y send the data s1, s2 respectively where $\mathrm{s}1 = \mathbf{H}1\mathbf{x}$ and $\mathrm{s}2 = \mathbf{H}2\mathbf{y}$ and $\mathbf{H}1, \mathbf{H}2$ are the parity check matrices of $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Without loss of generality let us suppose that the parity check matrices are in their systematic form. The decoding is done as follows:

$$\mathbf{x} = \mathbf{u}1\mathbf{G}1 + [\mathbf{0}|\mathrm{s}1], \mathbf{y} = \mathbf{u}2\mathbf{G}2 + [\mathbf{0}|\mathrm{s}2]. \tag{6}$$

Decoding involves finding a codeword $\mathbf{c}$ in $\mathcal{C}$ closest to $[\mathbf{0}|\mathrm{s}]$ where $[\mathbf{0}|\mathrm{s}] = [\mathbf{0}|\mathrm{s}1] + [\mathbf{0}|\mathrm{s}2]$. Thus $\mathbf{x}, \mathbf{y}$ can be recovered errorlessly. Q.E.D.

Let us consider the following example.
Example 2: Let $n = 7$ and $t = 1$ and $\mathcal{C}$ be the $(7, 4, 3)$ Hamming code. $\mathbf{G}1$ and $\mathbf{G}2$ are formed by partitioning the rows of $\mathbf{G}$, the generator matrix of $\mathcal{C}$. The encoders send the index of the coset of these subcodes containing the observed codevectors. After the reception of these indices, the decoder will decode the codevectors errorlessly.

Thus for the present case, any integral rate pairs in the Slepian-Wolf rate region can be achieved using linear codes. These concepts can be extended to codes on GF(q) and convolutional codes. The key idea is to build "linearly independent" subcodes of the main code under consideration (that captures the relevant correlation information) and assign them to different encoders. The encoders send the index of the coset of these subcodes containing the observed outcomes and the decoder needs to find from the cosets of the corresponding subcodes, a pair of codewords which are "closest" in the appropriate metric. Similar techniques can be developed for the encoding of sources with a fidelity criterion.

C Extension to encoding with a fidelity criterion:

These concepts can be extended to the encoding of correlated sources with a fidelity criterion[2]. The codes that we develop can be considered for the following two cases. In the former, there are two correlated sources that need to be encoded, and the decoder needs to get the best possible reconstruction of both of them. In the latter, the encoders observe a noisy version of a signal X, and encode their data such that the decoder gets an optimal reconstruction of X.

Asymmetric Encoding: In this setting, the encoders quantize their signals taking into account only the marginal distributions. One of the encoders sends its quantized data, and the other sends only a partial information (syndrome sequence) about the quantized data, thus reducing the rate of transmission by exploiting the correlation. This has been studied for the case of trellis codes in [7].

Symmetric Encoding: In this setting, both the encoders send only partial information (syndrome sequence) about their corresponding quantized data, thus reducing the rate of transmission by exploiting the correlation. This was first introduced and studied for the case of trellis codes and lattice codes in [8]. The signals are first remotely quantized using a root quantizer and two "linearly independent" subcodes of a main code are constructed and assigned to the two encoders. The main code is a subset of the product root quantizer reconstruction levels in N-dimensions. The encoders send the coset of their corresponding sub-codes in the root product quantizer. The decoder finds a pair of codewords from the cosets of the corresponding sub-codes which are closest in some metric. Finally, the decoder finds the optimal estimate of the signal of interest.

In this paper we consider a more general framework of generalized coset codes which encompasses trellis and lattice codes. Their distance properties are analyzed using a group-theoretic framework.

III. Generalized coset codes for distributed source coding

We consider the following formulation of the distributed source coding problem with applications to sensor networks. Let X be a random process observed by the sensors. The observations are characterized by the random processes Y_i for the i^{th} sensor. In this paper we restrict our attention to the case of two sensors. Let X be characterized as a sequence $\{X_i\}_{i=1}^{\infty}$ of independent identically distributed (*i.i.d.*) random variables. The observations $Y_1 = X + N_1$ and $Y_2 = X + N_2$, are characterized by the sequences $\{Y_{1i}\}_{i=1}^{\infty}$, $\{Y_{2i}\}_{i=1}^{\infty}$ where $\{N_{ji}\}_{i=1}^{\infty}$ (for j=1,2) are *i.i.d.* random variables which are independent of X. $Y_{1i} = X_i + N_{1i}$ and $Y_{2i} = X_i + N_{2i}$ where X_i,N_{ji} (for j=1,2) are modeled as zero-mean Gaussian random variables with variances given by 1 and σ^2 respectively. We need to encode Y_1 and Y_2 separately to be transmitted to the joint decoder whose task is to get the best estimate of X. We take the following approach for encoding. We quantize the observations by taking into account only the marginal distributions. The quantized codeword pair is referred to as the *active codeword pair.* Let S be the set of reconstruction levels of the root quantizer used in encoding. In this paper we consider the case of geometrically uniform S. Let $\mathbb{C}(S/S'; \mathcal{C})$ be a generalized coset code based on S, where the group code $\mathcal{C}$, the label group A, the partition S/S', the index set I, the mapping function $m : A \rightarrow S/S'$ and the label translate $\mathbb{C}(S/S'; \mathcal{C} + \mathbf{a})$ for $\mathbf{a} \in A^I$ are as defined in Section 2.

A Asymmetric Encoding:

It has been shown in [10] that $S^I/\mathbb{C}(S/S'; \mathcal{C})/(S')^I$ is a chain of geometrically uniform partitions and the label translates are congruent. Thus one of the encoders (say encoder-1) sends the index of the quantized codeword in S^I. The other encoder (encoder-2) sends the index of the label translate containing the quantized codeword in S^I. Let d_{min} be the minimum distance between any two codewords in $\mathbb{C}(S/S'; \mathcal{C})$. The decoder finds a codeword in the label translate whose index is received from the encoder-2, the codeword that is closest to the codeword received from the encoder-1. The decoder will recover the codewords correctly as long as the distance between the quantized codewords at the encoders is less than $d_{min}/2$.

[2] This problem is not the same as that of quantization followed by coding in the binary space. We need to construct codes with "good" distance properties directly in the Euclidean space. The analogy is similar to the power of coded modulation versus coding followed by modulation (see [14]).

B Symmetric Encoding:

This is the more interesting case where we need to maintain symmetry in the rates of transmission at the encoders. First we present the most general result which requires $U(S)$ to be Abelian. For the special cases [8] of trellis and lattice codes, which will be presented later, we do not require this restriction. We construct two subcodes of $\mathbb{C}(S/S';\mathcal{C})$ and assign them to the encoders. Let $\mathcal{C}_1$ be an Abelian subgroup (A is Abelian)of $\mathcal{C}$. Let the set of representatives of the cosets of $\mathcal{C}_1$ in $\mathcal{C}$ be given by $\mathbb{E} = \{\mathbf{e}_i : i \in I_E\}$ where the cardinality of $\mathbb{E}$ ($|\mathbb{E}|$) is equal to the order of $\mathcal{C}/\mathcal{C}_1$, and $I_E = \{1, 2, \ldots, |\mathbb{E}|\}$ and the set of representatives of $\mathcal{C}$ in A^I is given by $\mathbb{F} = \{\mathbf{a}_i : i \in I_F\}$ where $I_F = \{1, 2, \ldots, |\mathbb{F}|\}$. Thus we have

$$\bigcup_{\mathbf{e}\in\mathbb{E}} \bigcup_{\mathbf{c}\in\mathcal{C}_1} \mathbf{c} \oplus \mathbf{e} = \mathcal{C}, \tag{7}$$

where $\oplus$ denotes the group operation in A. Let $U(S'')$ be a normal subgroup of $U(S')$ which induces a partition S'/S''. The set of representatives of cosets of $U(S'')$ in $U(S')$ is given by $\mathbb{D} = \{u_{d_i} : i \in I_D\}$ where $I_D = \{1, 2, \ldots, |\mathbb{D}|\}$ and let the signal set associated with $\mathbb{D}$ be given by $S_{\mathbb{D}}$. Thus we have

$$\bigcup_{u'\in\mathbb{D}} \bigcup_{u\in U(S'')} u'[u(s_0)] = \bigcup_{u\in U(S')} u(s_0). \tag{8}$$

Example 3: Let S and $U(S)$ be as given in Example 1. Let $U(S') = T(2\mathbb{Z})$, and $I = \{1, 2, \ldots, 7\}$. Clearly S/S' is isomorphic to $\mathbb{Z}_2$. Let $\mathcal{C} \subset \mathbb{Z}_2^7$ be the (7,4,3) Hamming code with generator matrix $\mathbf{G}$ and let $\mathcal{C}_1$ is the set of codewords formed from the generator matrix $\mathbf{G}_1$ which is formed by taking any 2 rows of $\mathbf{G}$ (see Section 2). Now $|\mathbb{F}| = 8$, $|\mathbb{E}| = 4$ as there are 8 and 4 cosets of $\mathcal{C}$ in $A^7 = \mathbb{Z}_2^7$ and $\mathcal{C}_1$ in $\mathcal{C}$ respectively. Elements of $\mathbb{E}$ are the set of codewords formed by the generator matrix $\mathbf{G}_2$ which is formed by the remaining two rows of $\mathbf{G}$. Let $U(S'') = T(4\mathbb{Z})$. Thus $\mathbb{D} = \{T(0), T(2)\}$.

Using the above subgroup notation, we can express $\mathbb{C}(S/S';\mathcal{C})$ as

$$\mathbb{C}(S/S';\mathcal{C}) = \bigcup_{\mathbf{e}\in\mathbb{E}} \bigcup_{\mathbf{c}\in\mathcal{C}_1} \bigotimes_{k\in I} \bigcup_{u'\in\mathbb{D}} \bigcup_{u\in U(S'')} u_{e_k}\left[u_{c_k}\left[u'[u(s_0)]\right]\right]. \tag{9}$$

We construct two subcodes of $\mathbb{C}(S/S';\mathcal{C})$ where one of them is based on S/S'' and $\mathcal{C}_1$ and the other is based on $S/S_{\mathbb{D}}$ and $\mathbb{E}$. The subcodes of $\mathbb{C}(S/S';\mathcal{C})$ are defined as

$$\mathbb{C}_1(S/S'';\mathcal{C}_1) = \bigcup_{\mathbf{c}\in\mathcal{C}_1} u_{\mathbf{c}}\left((S'')^I\right) = \bigcup_{\mathbf{c}\in\mathcal{C}_1} \bigotimes_{k\in I} \bigcup_{u\in U(S'')} u_{c_k}[u(s_0)]; \tag{10}$$

$$\mathbb{C}_2(S/S_{\mathbb{D}};\mathbb{E}) = \bigcup_{\mathbf{e}\in\mathbb{E}} u_{\mathbf{e}}\left((S_{\mathbb{D}})^I\right) = \bigcup_{\mathbf{e}\in\mathbb{E}} \bigotimes_{k\in I} \bigcup_{u'\in\mathbb{D}} u_{e_k}[u'(s_0)]. \tag{11}$$

In some cases, $\mathbb{E}, \mathbb{D}$ are subgroups of $\mathcal{C}$ and $U(S')$ respectively. Clearly, the minimum distance of $\mathbb{C}_1(S/S'';\mathcal{C}_1)$ and $\mathbb{C}_2(S/S_{\mathbb{D}};\mathbb{E})$ is equal to d_{min} and they are subsets of S^I. These subcodes are assigned to two encoders. The encoders send the index of the cosets of these subcodes in S^I containing the quantized codeword. At the decoder, we need to find a pair of codewords from the cosets of the subcodes (as directed by the encoders), which are closest in distance. There is a finite probability of making a decoding error. The cosets of these subcodes are defined (for better notation) as follows

$$\mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a}) = u_{\mathbf{a}}[u_{\mathbf{f}}[u_{\mathbf{t}}\{\mathbb{C}_1(S/S'';\mathcal{C}_1)\}]]$$

$$= \bigcup_{\mathbf{c}\in\mathcal{C}_1} \bigotimes_{k\in I} \bigcup_{u\in U(S'')} u_{a_k}\left[u_{f_k}\left[u_{c_k}[u_{t_k}[u(s_0)]]\right]\right], \tag{12}$$

$$\mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b}) = u_{\mathbf{b}}[u_{\mathbf{h}}[u_{\mathbf{g}}\{\mathbb{C}_2(S/S_{\mathbb{D}};\mathbb{E})\}]]$$

$$= \bigcup_{\mathbf{e}\in\mathbb{E}} \bigotimes_{k\in I} \bigcup_{u'\in\mathbb{D}} u_{b_k}\left[u_{e_k}\left[u_{h_k}[u'[u_{g_k}(s_0)]]\right]\right], \tag{13}$$

for any $\mathbf{f} \in \mathbb{E}$, $u_{\mathbf{t}} \in (\mathbb{D})^I$, $\mathbf{a}, \mathbf{b} \in (\mathbb{F})^I$, $\mathbf{h} \in \mathcal{C}_1$ and $u_{\mathbf{g}} \in (U(S''))^I$. In other words, for $\mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$, $\mathbf{f}$ and $u_{\mathbf{t}}$ specify its cosets in $\mathbb{C}_1(S/S';\mathcal{C}+\mathbf{a})$ where $\mathbf{a}$ specifies the coset of $\mathbb{C}_1(S/S';\mathcal{C})$ in S^I. Similarly, it is true for the other subcode.

For a precise characterization of the distance properties of these subcodes, we need the following definition. For any pair of cosets $\mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$, $\mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$, define [8] $\alpha(\mathbf{f},\mathbf{t},\mathbf{a},\mathbf{h},\mathbf{g},\mathbf{b})$ = minimum of distances $\|\mathbf{r}_1 - \mathbf{r}_2\|$ between any two codewords $\mathbf{r}_1 \in \mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$ and $\mathbf{r}_2 \in \mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$ such that $\exists$ at least one pair $(\mathbf{r}_3,\mathbf{r}_4) \neq (\mathbf{r}_1,\mathbf{r}_2)$ such that $\|\mathbf{r}_3 - \mathbf{r}_4\| \leq \|\mathbf{r}_1 - \mathbf{r}_2\|$, $\mathbf{r}_3 \in \mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$ and $\mathbf{r}_4 \in \mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$ (*i.e.*, $(\mathbf{r}_1,\mathbf{r}_2)$ *is the pair with second smallest distance*). For the subcodes defined as before, we define the correlation distance d_c, as follows:

$$d_c = \underset{\mathbf{f}\in\mathbb{E},\, u_{\mathbf{t}}\in(\mathbb{D})^I,\, \mathbf{a},\mathbf{b}\in(\mathbb{F})^I,\, \mathbf{h}\in\mathcal{C}_1,\, u_{\mathbf{g}}\in(U(S''))^I}{\text{minimum}} \{\alpha(\mathbf{f},\mathbf{t},\mathbf{a},\mathbf{h},\mathbf{g},\mathbf{b})\} \tag{14}$$

It can be noted that for the decoder to make an error, the distance between the quantized codewords must be at least d_c. The distance properties are given by the following theorem.

Theorem: For the above partition of $\mathbb{C}(S/S';\mathcal{C})$, $d_c \geq d_{min}/2$.

Proof: Let $\mathbf{w}_1 \in \mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$ and $\mathbf{w}_2 \in \mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$ be an active codeword pair. And let $\|\mathbf{w}_1 - \mathbf{w}_2\| < d_{min}/2$. Consider any $\mathbf{w}_1' \in \mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$, $\mathbf{w}_1 \neq \mathbf{w}_1'$. This can be represented as

$$\mathbf{w}_1' = u_{\mathbf{l}}[\tilde{u}[\mathbf{w}_1]], \tag{15}$$

for some $\mathbf{l} \in \mathcal{C}_1$ and $\tilde{u} \in (U(S''))^I$. Let

$$\mathbf{w}_2' = u_{\mathbf{l}}[\tilde{u}[\mathbf{w}_2]]. \tag{16}$$

Clearly, $\|\mathbf{w}_1 - \mathbf{w}_2\| = \|\mathbf{w}_1' - \mathbf{w}_2'\| < d_{min}/2$, $\mathbf{w}_2' \notin \mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$ and $\mathbf{w}_2, \mathbf{w}_2' \in \mathbb{C}(S/S';\mathcal{C}+\mathbf{b})$. Since $\mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b}) \subset \mathbb{C}(S/S';\mathcal{C}+\mathbf{b})$, there can not exist another pair $(\mathbf{w}_3,\mathbf{w}_4) \neq (\mathbf{w}_1,\mathbf{w}_2)$ such that $\|\mathbf{w}_3 - \mathbf{w}_4\| < d_{min}/2$, $\mathbf{w}_3 \in \mathbb{BC}_1(\mathbf{f},\mathbf{t},\mathbf{a})$ and $\mathbf{w}_4 \in \mathbb{BC}_2(\mathbf{h},\mathbf{g},\mathbf{b})$. Q.E.D.

The decoding can be done in a manner that is similar to the case of asymmetric encoding with a slight modification of the decoding algorithm of generalized coset codes. Due to lack of space we are not presenting the details of decoding and the computational complexity in this paper. In [8], decoding algorithms for the special cases of simple trellis codes and lattice codes were given. For more general decoding algorithms for coset codes with more than two users, the readers are referred to [17]. Thus the performance of symmetric encoding in terms of minimum distance, is not compromised with respect to that of asymmetric encoding. We partition both $\mathcal{C}$ and $U(S')$ (subgroups of A^I and $U(S)$ respectively) to obtain the subcodes. When these partitions are trivial, we get the special cases of subcodes of trellis and lattice codes as given in [8]. The subcodes are constructed by obtaining the subgroups of $\mathcal{C}$ and $U(S')$ such that one of the subcodes $\mathbb{C}_1(S/S'';\mathcal{C}_1)$ is based on S'' and $\mathcal{C}_1$ and the other $\mathbb{C}_2(S/S_{\mathbb{D}};\mathbb{E})$ is based on $S_{\mathbb{D}}$ and $\mathbb{E}$. We consider the following special cases of particular interest.

IV. Special Cases

In this section we consider the special cases where one of the subgroups $\mathcal{C}_1$ and $\mathbb{D}$ is trivial. We give examples of trellis codes and lattice codes to illustrate the concepts.

A Case 1: $\mathcal{C}_1 = \mathcal{C}$ and $\mathbb{D} = U(S')$:

In this case we consider the trivial subgroups where $\mathcal{C}_1 = \mathcal{C}$ and $\mathbb{D} = U(S')$. Thus the subcodes of $\mathbb{C}(S/S';\mathcal{C})$ are given by

$$\mathbb{BC}_1(\mathbf{t},\mathbf{a}) = u\mathrm{a}\left[u\mathrm{t}\left\{\bigcup_{C\in\mathcal{C}}\bigotimes_{k\in I} u_{c_k}[s_0]\right\}\right], \tag{17}$$

$$\mathbb{BC}_2(\mathbf{h},\mathbf{b}) = u\mathrm{b}\left[u\mathrm{h}\left\{\bigotimes_{k\in I}\bigcup_{u'\in U(S')}[u'(s_0)]\right\}\right], \tag{18}$$

where $u\mathrm{t} \in (U(S'))^I$, $\mathbf{a},\mathbf{b} \in (\mathbb{F})^I$, $\mathbf{h} \in \mathcal{C}$. For this case it can be shown that $d_c \geq d_{min}/2$ even when the generating group $U(S)$ is not Abelian.

Example 4: Trellis codes built on lattices and cosets [14, 15, 9] can be formulated as generalized coset codes with S being a translate of a lattice, Λ, and is represented as $\mathbb{C}(\Lambda/\Lambda';\mathcal{C})$, where Λ' is the sublattice of Λ. We give the following example to illustrate the concepts:

Let $\Lambda = \mathbb{Z} + 1/2$ and let the generating group be $U(\Lambda) = (V.T(2\mathbb{Z}))$. Let the normal subgroup $U(\Lambda')$ of $U(\Lambda)$ be given by: $U(\Lambda') = (T(4\mathbb{Z}))$. Clearly Λ/Λ' is isomorphic to $\mathbb{Z}_2^2$. Let $\mathcal{C}$ be a rate 1/2 convolutional code with the generator polynomial matrix given by $\mathbf{G}(\mathbf{t}) = [1\ \frac{t}{1+t^2}]$. The subcodes are shown in Fig. 4 and Fig. 5.

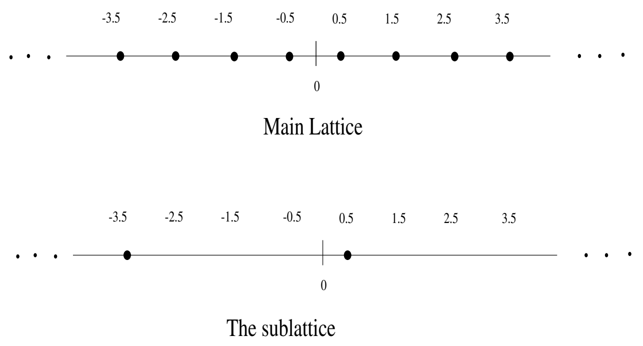

Figure 4: Lattice Λ and the sublattice Λ'. This sublattice is assigned to one of the encoders.

B Case 2: $\mathcal{C} = A^I$ and $I = \{1\}$:

Lattice codes [16] can be trivially obtained when $I = \{1\}$. Even for this case, the generating group, $U(\Lambda)$ need not be Abelian. Let $U(\Lambda'')$ be the normal subgroup of $U(\Lambda')$ and let $\mathbb{D}$ represents the set of coset representatives as defined before. The subcodes can be obtained as follows:

$$\mathbb{C}_1 = \bigcup_{u\in U(\Lambda'')} u(s_0), \mathbb{C}_2 = \bigcup_{u'\in\mathbb{D}} u'(s_0). \tag{19}$$

As before, it can be shown that $d_c \geq d_{min}/2$. The decoding complexity is the same as that of decoding a codeword in Λ'. We give three illustrative examples to demonstrate the concepts.

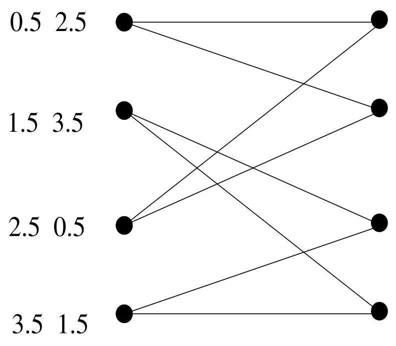

Figure 5: The trellis associated with the second subcode. The generator matrix polynomial is given by $\mathbf{G}(\mathbf{t}) = [1\ \frac{t}{1+t^2}]$. This trellis code is assigned to the other encoder.

Example 5: In this example, let $\Lambda = \mathbb{Z}$ with the generating group given by $T(\mathbb{Z})$. Let the subgroups $U(\Lambda')$ and $U(\Lambda'')$ be given by $T(3\mathbb{Z})$ and $T(3M\mathbb{Z})$ respectively for some $M \in \mathbb{Z}$. The d_{min} of Λ' is 3. The set of coset representatives of $U(\Lambda'')$ in $U(\Lambda')$ is given by $\mathbb{D} = \{T(3i) : i \in \{0, 1, \ldots, M-1\}\}$ and its associated signal set is given by $\Lambda_{\mathbb{D}} = \{3i : i \in \{0, 1, \ldots, M-1\}\}$. Each encoder quantizes the observation using Λ. The encoder-1 sends the index of the coset of $U(\Lambda'')$ in $U(\Lambda)$, and the encoder-2 sends the index of the coset of $\Lambda_{\mathbb{D}}$ in Λ. By varying M, we can balance the rates between two encoders with a performance measure $d_c \geq 1.5$. This example is shown in Fig. 6. It can be noted that $\mathbb{D}$ is not a lattice.

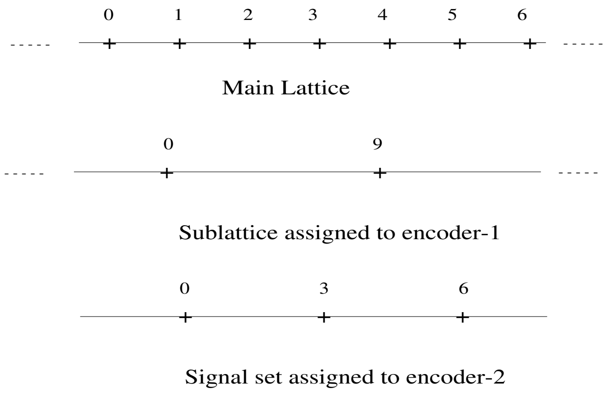

Figure 6: The lattice $\mathbb{Z}$: For M=3, the sublattice $9\mathbb{Z}$ is assigned to encoder-1. Also shown is the signal set $\{0, 3, 6\}$ assigned to the encoder-2.

Example 6: Let Λ be $\mathbb{Z}^2$ with a generating group given by $T(\mathbb{Z}^2)$. The normal subgroups $U(\Lambda')$ and $U(\Lambda'')$ are given by $T(3\mathbb{Z}^2)$ and $T(3\mathbb{Z},0)$ respectively. The subcodes associated with the encoders are given by $(3\mathbb{Z},\mathbf{0})$ and $(\mathbf{0},3\mathbb{Z})$. This is shown in Fig. 7. The encoders send the index of the cosets of their corresponding subcodes in $\mathbb{Z}^2$, containing the active codeword pair independently. The correlation distance, $d_c \geq$

1.5. It can be noted that in this case, $\mathbb{D}$ is a lattice.

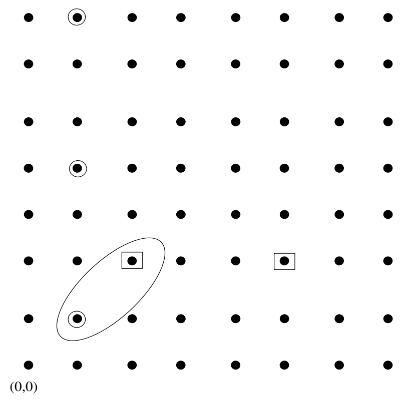

Figure 7: An example of a lattice $\mathbb{Z}^2$. The sublattices associated with the encoders are given by $(3\mathbb{Z}, \mathbf{0})$ and $(\mathbf{0}, 3\mathbb{Z})$. The encoders send the index of the coset of these sublattices in $\mathbb{Z}^2$ containing the quantized codewords. There exists a unique pair as shown such that the distance between them is minimum among all the other pairs.

Example 7: Let Λ be $(\frac{\sqrt{3}}{2}, \frac{1}{2})\mathbb{Z} + (\frac{\sqrt{3}}{2}, -\frac{1}{2})\mathbb{Z}$ as shown in Fig. 8. The sublattices associated with the encoders are also shown.

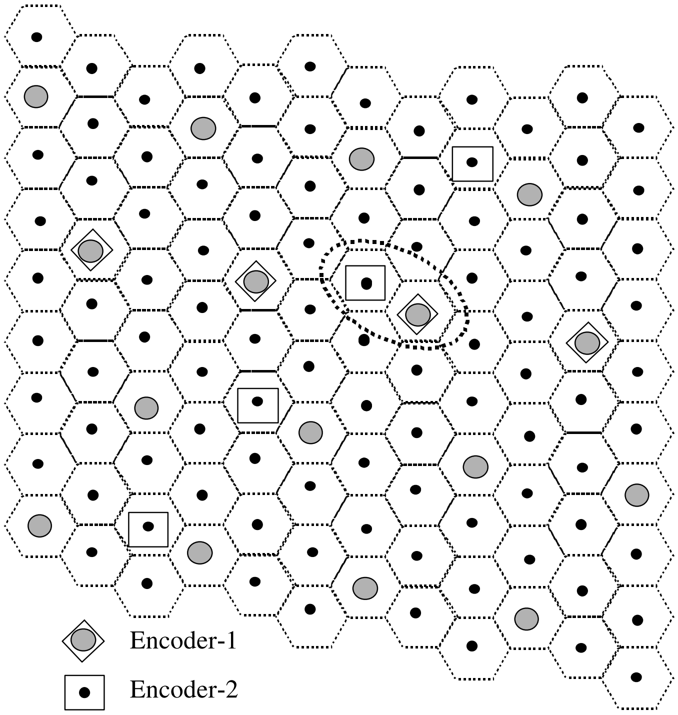

Figure 8: The lattice considered is $(\frac{\sqrt{3}}{2}, \frac{1}{2})\mathbb{Z} + (\frac{\sqrt{3}}{2}, -\frac{1}{2})\mathbb{Z}$. The sublattice Λ' is shown with big shaded circles, which is given by $(\frac{3\sqrt{3}}{2}, -\frac{1}{2})\mathbb{Z} + (\sqrt{3}, 2)\mathbb{Z}$. The sublattices associated with the encoders are $(\frac{3\sqrt{3}}{2}, -\frac{1}{2})\mathbb{Z}$ and $(\sqrt{3}, 2)\mathbb{Z}$. The encoders send the index of the cosets of these sublattices containing the quantized codewords in Λ.

V. Conclusions

In this paper we presented a formulation of generalized coset codes in a group-theoretic setting for symmetric and asymmetric of distributed source coding. It is also shown that the performance of symmetric encoding matches that of asymmetric encoding with the same computational complexity. This formulation is very general and encompasses the class of codes given by lattice codes and trellis codes on lattices and cosets. We have presented illustrative examples of trellis codes and lattice codes.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory.* New York: Wiley, 1991.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, January 1976.

[4] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, 1978.

[5] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, November 1987.

[6] T. Berger, "Multiterminal source coding," *in The information theory approach to communication, (CISM Courses and Lecture Notes No. 229), G. Longo, Ed., Wien and New York: Springer-Verlag.*, 1977.

[7] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS) : Design and construction," *Proc. IEEE Data Compression Conference, Snowbird, UT*, March 1999.

[8] S. S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor networks," *Proc. IEEE Data Compression Conference, Snowbird, UT*, March 2000.

[9] G. D. Forney, "Coset codes- part 1: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, September 1988.

[10] G. D. Forney, "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, September 1991.

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

[12] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes.* Elsevier-North-Holland, 1977.

[13] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2–10, January 1974.

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

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

[16] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups.* New York: Springer-Verlag, 1988.

[17] S. S. Pradhan and K. Ramchandran, "Decoding algorithms for distributed source coding with multiple encoders for generalized coset codes," *preprint* , Jan, 2000.