Lattice Codes Achieve the Capacity of Common Message Gaussian Broadcast Channels With Coded Side Information

Lakshmi Natarajan, Yi Hong, *Senior Member, IEEE*, and Emanuele Viterbo, *Fellow, IEEE*

Abstract—Lattices possess elegant mathematical properties which have been previously used in the literature to show that structured codes can be efficient in a variety of communication scenarios, including coding for the additive white Gaussian noise channel, dirty-paper channel, Wyner–Ziv coding, coding for relay networks, and so forth. We consider the family of single-transmitter multiple-receiver Gaussian channels, where the source transmits a set of common messages to all the receivers (multicast scenario), and each receiver has *coded side information*, i.e., prior information in the form of linear combinations of the messages. This channel model is motivated by applications to multi-terminal networks, where the nodes may have access to coded versions of the messages from previous signal hops or through orthogonal channels. The capacity of this channel is known and follows from the work of Tuncel (2006), which is based on random coding arguments. In this paper, following the approach of Erez and Zamir, we design lattice codes for this family of channels when the source messages are symbols from a finite field $\mathbb{F}_p$ of prime size. Our coding scheme utilizes Construction A lattices designed over the same prime field $\mathbb{F}_p$, and uses *algebraic binning* at the decoders to expurgate the channel code and obtain good lattice subcodes, for every possible set of linear combinations available as side information. The achievable rate of our coding scheme is a function of the size p of underlying prime field, and approaches the capacity as p tends to infinity.

Index Terms—Capacity, Construction A, Gaussian broadcast channel, lattice, multicast, side information, structured codes.

I. INTRODUCTION

INFORMATION-THEORETIC results often rely on random coding arguments to prove the existence of good codes. Usually, the codebook is constructed by randomly choosing the components of each codeword independently and identically from a judiciously chosen probability distribution. While this technique is powerful, the resulting codebooks do not exhibit any structure that may be of practical interest. One such desirable structure is linearity, which allows complexity reductions at the encoder and decoder by utilizing efficient algebraic processing techniques. Further, in certain communication scenarios, coding schemes based on linear codes yield a larger achievable rate region than random code ensembles, as was shown by Körner and Marton [1] for a distributed source coding problem. Structured coding schemes have been widely studied in the literature, especially for communications in the presence of side information and in multi-terminal networks. For an overview of structured coding schemes we refer the reader to [2] and [3] and the references therein.

For communication in the wireless domain, structured codes can be obtained by choosing finite subsets of points from lattices [2], [4]–[6]. A lattice is an infinite discrete set of points in the Euclidean space that are regularly arranged and are closed under addition. Codes based on lattices, known as *(nested) lattice codes* or *Voronoi codes*, are the analogues of linear codes in wireless communications. Efficient lattice based strategies are known for a variety of communication scenarios, such as for achieving the capacity of the point-to-point additive white Gaussian noise (AWGN) channel [7]–[11], for dirty-paper coding [2], [12], the Wyner–Ziv problem [2] and communication in relay networks [13]–[16], to name only a few.

In this paper we present good lattice strategies for communication in common message Gaussian broadcast channels, which we refer to as the *multicast channel*, where receivers have prior side information about the messages being transmitted. In particular, we assume that the transmitter is multicasting K message symbols $w_1, \ldots, w_K$ from a finite field $\mathbb{F}_p$, of prime size p, to all the receivers, and each receiver may have *coded side information* about the messages: the prior knowledge of the values of (possibly multiple) $\mathbb{F}_p$-linear combinations of $w_1, \ldots, w_K$. The number of linear combinations available as side information and the coefficients of these linear combinations can differ from one receiver to the next. The capacity of this channel is known and follows from the results of Tuncel [17], where the achievability part utilizes an ensemble of codebooks generated using the Gaussian distribution.

The multiuser channel considered in this paper is a noisy version of a simple special case of *index coding* [18]–[20]. The index coding problem considers a *noiseless* broadcast link where each receiver demands a subset of the source messages

and knows the values of some other subset as side information. A generalization of the index coding problem in which the receivers have access to linear combinations of messages was studied recently in [21] and [22]. The specific instance of index coding where each receiver demands all the messages from the source corresponds to a noiseless multicast channel and has a simple optimum solution based on maximum distance separable (MDS) codes [23]. When the channel is noisy, capacity-achieving coding schemes based on structured codes are not available. In this paper we design lattice-based strategies for multicasting over the AWGN channel where the side information at the receivers is in the form of linear combinations of source messages.

The case of Gaussian multicast channel with coded side information is motivated by applications to multi-terminal communication networks. It is known that signal interference in wireless channels can be harnessed by decoding linear combinations of transmit messages instead of either treating interference as noise or decoding interference along with the intended message [14], [15]. When such a technique is used in a mutli-hop communication protocol, one encounters receivers that have coded side information obtained from transmissions in the previous phases. Similarly, in a network that consists of both wired and wireless channels, the symbols received from wired links can be utilized as side information for decoding the wireless signals. If a linear network code is used in the wired part of the network, then the side information is in the form of linear combinations of the source messages.

Example 1 (Communciation in Relay Networks): Consider a wireless network with two base stations BS_1 and BS_2, that hold message symbols w_1 and w_2, respectively. The base stations are required to multicast w_1 and w_2 to four user nodes $U_1, \ldots, U_4$ through the relay node R, see Fig. 1. In the first phase of the protocol, BS_1 and BS_2 encode the data symbols w_1 and w_2, and transmit the resulting codewords simultaneously. By using the decoding technique of compute-and-forward [15], U_3 reliably decodes some linear combination $s_1 w_1 + s_2 w_2$, $s_1, s_2 \in \mathbb{F}_p$, from the received noisy superposition of the two transmit signals. On the other hand, R has a higher signal-to-noise ratio and successfully decodes both w_1 and w_2 by behaving as a multiple-access receiver. Further, there is no signal interference at U_1 and U_2, and these two nodes reliably decode w_1 and w_2, respectively.

We observe that the second phase of the protocol is a common message broadcast channel with coded side information at the receivers: the relay needs to multicast w_1, w_2 to four user nodes, the first three users U_1, U_2, U_3 have prior knowledge of the linear combinations $w_1 + 0w_2$, $0w_1 + w_2$ and $s_1 w_1 + s_2 w_2$, respectively, while the fourth user has no such side information. ∎

Example 2 (Wireless Overlay for Wired Networks): Assume a network of noiseless wired links in the form of a directed acyclic graph, where the source node v_s desires to multicast K independent messages $w_1, \ldots, w_K \in \mathbb{F}_p$ to a set of destination nodes $\mathcal{D}$. The wireline network employs a traditional *(scalar) linear network code* [24]–[26], i.e., the symbol transmitted on each outgoing edge of a node is an element of $\mathbb{F}_p$ generated as a linear combination of the symbols received
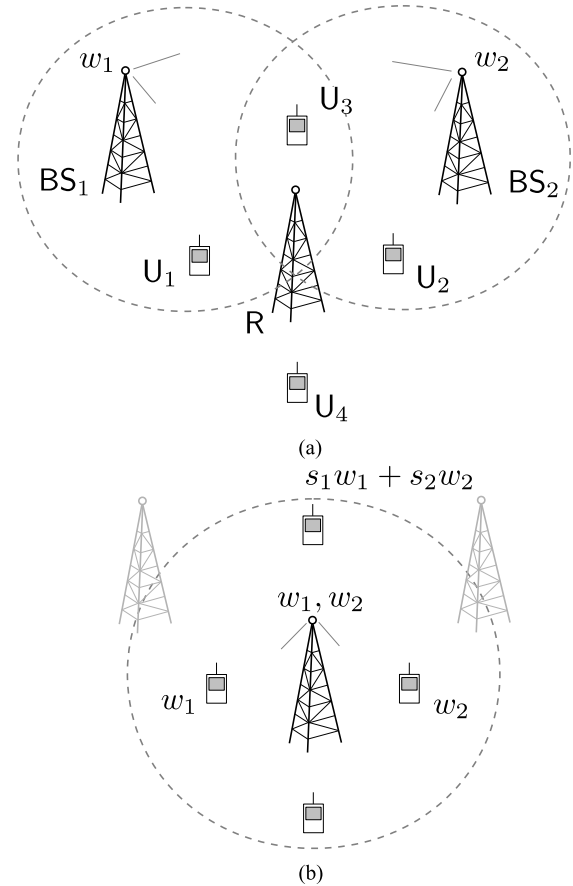

Fig. 1. A relay network where one encounters a common message broadcast channel with coded side information at the receivers. (a) Multiple-access phase: The relay R decodes both w_1 and w_2, while U_1, U_2, U_3 decode w_1, w_2 and $s_1 w_1 + s_2 w_2$, respectively. (b) Multicast phase: R multicasts w_1 and w_2 to all four user nodes. Three of the users have the knowledge of some linear combination of w_1 and w_2, while the fourth user has no side information.

on its incoming edges. At every destination node $v_d \in \mathcal{D}$, the decoder attempts to recover the K message symbols from their $\mathbb{F}_p$-linear combinations received on its incoming edges. Recovery is possible if and only if the number of linearly independent equations available at v_d is K. It is known that the maximum number of linearly-independent equations that can be made available at v_d is $\min\{\text{max-flow}(v_d), K\}$, where max-flow$(v_d)$ is the maximum number of edge-disjoint paths from v_s to v_d, see [26]. It follows that multicasting is possible if and only if max-flow$(v_d) \geq K$ for every $v_d \in \mathcal{D}$.

Now suppose there exist destination nodes with max-flow less than K, i.e., the communication demands are beyond the wireline network's capacity. A solution to meet the demands is to broadcast a wireless signal from the source to fill the capacity deficiency of the wired network, see Fig. 2. At each destination, the $\mathbb{F}_p$-linear combinations obtained from the wireline network serve as side information to decode the wireless broadcast signal. ∎

A special case of coded side information is the Gaussian multicast channel where each receiver has prior knowledge of the values of some subset of the K messages. The known capacity-achieving coding schemes for this special case are based on random coding using i.i.d. (independent and identically distributed) codewords [17], [27]–[31]. Existence of
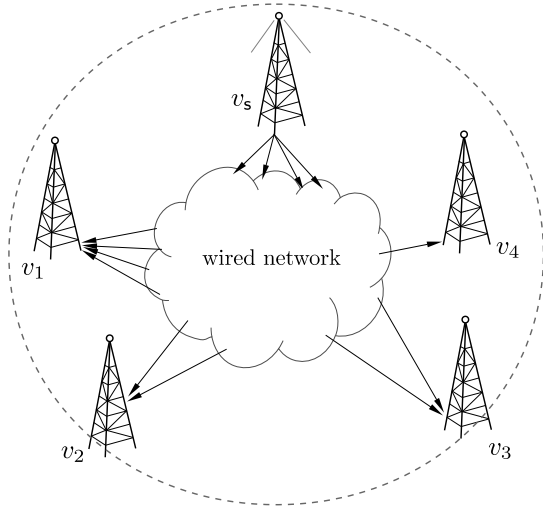
Fig. 2. Wireless multicast as an overlay for a wired network: the source node v_{s} encodes $w_1, \ldots, w_K$ and broadcasts wirelessly to the destination nodes $v_1, \ldots, v_4$ in order to supplement the communication through the wired network. Each destination receives (possibly multiple) $\mathbb{F}_p$-linear combinations of the source symbols through a linear (wireline) network code and uses this information to decode the wireless signal broadcast by v_{s}.

lattice based capacity-achieving coding schemes were proved in [28] and [32] for the special case where the number of messages and receivers are two and each receiver has the knowledge of one of the messages. Constructions of binary codes for this channel were proposed in [33]–[35]. Explicit constructions of lattice codes were given in [36] and [37] that convert receiver side information into additional apparent coding gain in the AWGN channel. Codes based on quadrature amplitude modulation were constructed in [38] and [39]. In [32], explicit codes based on lattices and coded modulation have been designed that perform within a few decibels of capacity when the number of receivers is two and each knows one of the two messages being transmitted.

The objective of this paper is to prove that lattice codes can achieve the capacity of the common message Gaussian broadcast channels with coded side information. We use the information-theoretic framework set by Erez and Zamir [8] to this end. The proposed coding scheme uses lattices obtained by applying Construction A to linear codes over the prime field $\mathbb{F}_p$ which is the alphabet of the source messages. The achievable rate of our lattice-based coding scheme is a function of the prime p, and approaches the capacity of the common message Gaussian broadcast channel as $p \to \infty$.

Our decoding scheme involves *algebraic binning* [2] where the receiver side information is used to expurgate the channel code and obtain a lower rate subcode. The set of linear equations available as side information may differ from one receiver to another, and hence, each receiver must employ a different binning scheme for the same channel code. The coding scheme ensures that the binning performed at each receiver produces a good lattice subcode of the transmitted code. Following expurgation, each receiver decodes the channel output by minimum mean square error (MMSE) scaling and quantization to an infinite lattice. The algebraic structure of the coding scheme facilitates the performance analysis by decomposing the original channel into multiple independent

point-to-point AWGN channels – one corresponding to each receiver – where each of the point-to-point AWGN channels uses a lattice code for communication. Unlike [8], where achievability in a point-to-point AWGN channel was proved using error exponent analysis, we provide a direct proof based only on simple counting arguments.

As a corollary to the main result, we obtain an alternative proof of the goodness of lattice codes in achieving the capacity of the point-to-point AWGN channel. Previous proofs of this result presented in [8]–[10] also use ensembles of lattices obtained by applying Construction A to random linear codes over a prime field $\mathbb{F}_p$; see also [40], [41]. While [8] used primes p that were exponential in the code length n, [9] and [10] improved this result to let p grow as $n^{1.5}$ and $n^{0.5}$, respectively. The corollary presented in this paper further improves these results by enabling a choice of the prime p which is independent of the code length n but is a function only of the gap between the desired rate and the channel capacity.

Lattices have been used to design powerful physical-layer coding schemes for wireless networks consisting of multiple sources, relays and destinations [13]–[16]. In these networks information from the source nodes is conveyed to the destination nodes through relays over multiple hops and time slots. In each time slot, a set of nodes act as transmitters and every other node in their range observes a linear superposition of the transmitted signals perturbed by AWGN. Lattice coding schemes for these networks are designed such that each receiver can reliably decode the observed noisy superposition to a linear combination of source messages which it then proceeds to transmit in the next time slot. Every destination node decodes its desired messages once it collects sufficiently many linear combinations. In contrast, in this paper we consider a single hop interference-free transmission in a multicast channel consisting of one transmitter and multiple destination nodes that are aided by coded side information. Our objective is to design coding schemes that can utilize prior knowledge at these receivers rather than exploit wireless interference arising from multiple simultaneous transmissions, as often experienced in relay networks.

The organization of this paper is as follows. We introduce the channel model in Section II-A and review the relevant background on lattices and lattice codes in Section II-B. In Section III, we state the main theorem, and describe the lattice code ensemble and encoding and decoding procedures. We prove the main theorem and state a few corollaries in Section IV, and finally, we discuss some concluding remarks in Section V.

Notation: Matrices and column vectors are denoted by bold upper and lower case letters, respectively. The symbol $\| \cdot \|$ denotes the Euclidean norm of a vector, and $(\cdot)^{\mathsf{T}}$ is the transpose of a matrix or a vector. The Kronecker product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is $\boldsymbol{A} \otimes \boldsymbol{B}$, $\boldsymbol{I}_{\ell}$ is the $\ell \times \ell$ identity matrix, and $\boldsymbol{0}$ is the all zero matrix of appropriate dimension. The symbol $\log(\cdot)$ denotes logarithm to the base 2 and $\ln(\cdot)$ denotes logarithm to the base e. The expectation operator is denoted by $\mathbb{E}$. The symbol $\mathcal{M} \backslash \mathcal{N}$ denotes the elements in the set $\mathcal{M}$ that do not belong to the set $\mathcal{N}$.

II. CHANNEL MODEL AND LATTICE PRELIMINARIES

A. Channel Model and Problem Statement

We consider a (non-fading) common message Gaussian broadcast channel with a single transmitter and finitely many receivers, where all terminals are equipped with single antennas. The transmitter operates under an average power constraint and the receivers are affected by additive white Gaussian noise with possibly different noise powers. There are K independent messages $w_1, \ldots, w_K$ at the transmitter that assume values with a uniform probability distribution from a prime finite field $\mathbb{F}_p$. Each receiver desires to decode all the K messages while having prior knowledge of the values of some $\mathbb{F}_p$-linear combinations of the messages $w_1, \ldots, w_K$. Consider a generic receiver that has access to the values u_m, $m = 1, \ldots, M$, of the following set of M linear equations

$$\sum_{k=1}^{K} s_{m,k} w_k = u_m, \quad m = 1, \ldots, M.$$

We will denote this side information configuration using the matrix $S = [s_{m,k}] \in \mathbb{F}_p^{M \times K}$, where each row of S represents one linear equation. Any row of S that is linearly dependent on the other rows represents redundant information and can be discarded with no loss to the receiver side information, and hence, with no loss to system performance. Hence, without loss in generality, we will assume that the rows of S are linearly independent over $\mathbb{F}_p$, i.e., $\mathrm{rank}(S) = M$, and $M < K$. Note that the values of S and M can be different across the receivers. A receiver with no side information is represented with an empty matrix for S (with $M = 0$).

A receiver in the multicast channel is completely characterized by its *(coded) side information matrix* S and the variance σ^2 of the additive noise. If we assume that the average transmit power at the source is 1, then the signal-to-noise ratio at this receiver is $\mathsf{SNR} = \frac{1}{\sigma^2}$. We will denote a receiver by the pair (S, σ^2), where S is any matrix over $\mathbb{F}_p$ with K columns and linearly independent rows, and $\sigma^2 > 0$. Note that *uncoded* side information, i.e., the prior knowledge of the values of a size M subset of $w_1, \ldots, w_K$, is a special case, and hence, is contained within the definition of our channel model.

Example 3: Consider a source transmitting $K = 3$ symbols, w_1, w_2, w_3, from the finite field $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$. A receiver that has prior knowledge of the value of w_2 has side information matrix $S = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$. This corresponds to the equation $0w_1 + 1w_2 + 0w_3$, and the number of linearly independent equations at this receiver is $M = \mathrm{rank}(S) = 1$.

Now consider another receiver that has the knowledge of the values of the following three equations: $w_1 + 4w_2 + 3w_3$, $4w_1 + 3w_2$ and $2w_1 + w_2 + 3w_3$. In matrix form, this side information is represented by

$$\begin{pmatrix} 1 & 4 & 3 \\ 4 & 3 & 0 \\ 2 & 1 & 3 \end{pmatrix},$$

where the three rows represent the three equations, in that order. The first row of this matrix is equal to the sum (over $\mathbb{F}_5$) of the second and third rows, and hence, the side information from the first equation is redundant and can be discarded. Since the remaining two rows are linearly independent, the side information at this receiver can be represented by the following matrix that consists of these two rows,

$$S = \begin{pmatrix} 4 & 3 & 0 \\ 2 & 1 & 3 \end{pmatrix}.$$

The number of linearly independent equations at this receiver is $M = \mathrm{rank}(S) = 2$. ∎

From elementary linear algebra we know that if the values u_m of M linearly independent combinations of the variables $w_1, \ldots, w_K$ are given, then the set of all possible solutions of $(w_1, \ldots, w_K)$ is a coset of a $(K - M)$ dimensional linear subspace of $\mathbb{F}_p^K$. Since the a priori probability distribution of $w_1, \ldots, w_K$ is uniform, we conclude that, given the side information values u_m, $m = 1, \ldots, M$, the probability distribution of $(w_1, \ldots, w_K)$ is uniform over this coset. Using the fact that the number of elements in the coset is p^{K-M}, we observe that the conditional entropy of $(w_1, \ldots, w_K)$ given the side information is

$$H(w_1, \ldots, w_K \mid u_1, \ldots, u_M) = \log\left(p^{K-M}\right)$$
$$= (K - M) \log p. \quad (1)$$

Suppose we want to transmit, on the average, one realization of $(w_1, \ldots, w_K)$ in every κ uses of the broadcast channel. The transmission rate of each message is $R = \frac{1}{\kappa} \log p$ b/dim (bits per real dimension or bits per real channel use).

For the simplicity of exposition, we consider only the symmetric case where all the K messages are required to be transmitted at the same rate R. The general scenario, where the messages are of different rates, can be reduced to the symmetric case through *rate-splitting*: if there are K' messages with transmission rates $r_1, \ldots, r_{K'}$, respectively, then by splitting each of these original sources into multiple virtual sources, one can generate a set of K sources $(K \geq K')$ such that their rates $R_1, \ldots, R_K$ are as close to each other as required.

We will assume that the encoding at the transmitter is performed on a block of ℓ independent realizations of the K message symbols, i.e., the source jointly encodes K message vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K \in \mathbb{F}_p^\ell$. The transmitter uses an n-dimensional channel code $\mathcal{X} \subset \mathbb{R}^n$ together with a function

$$\rho : \mathbb{F}_p^\ell \times \cdots \times \mathbb{F}_p^\ell \to \mathcal{X}$$

to jointly encode the K message vectors. The number of codewords in $\mathcal{X}$ is $p^{K\ell}$, and we will assume that the codebook $\mathcal{X}$ satisfies the per-codeword power constraint

$$\frac{\|\boldsymbol{x}\|^2}{n} \leq 1, \quad \text{for all } \boldsymbol{x} \in \mathcal{X}. \quad (2)$$

The average number of channel uses to transmit each realization of $(w_1, \ldots, w_K)$ is $\kappa = \frac{n}{\ell}$. The resulting rate of transmission of each of the K messages is

$$\frac{\log\left(p^\ell\right)}{n} = \frac{\ell}{n} \log p \text{ b/dim.}$$

The sum rate of all the messages is $\frac{K\ell}{n} \log p$ b/dim.

The side information at (S, σ^2) over a block of ℓ realizations of the K message symbols is of the form $\sum_{k=1}^{K} s_{m,k} \boldsymbol{w}_k = \boldsymbol{u}_m$,
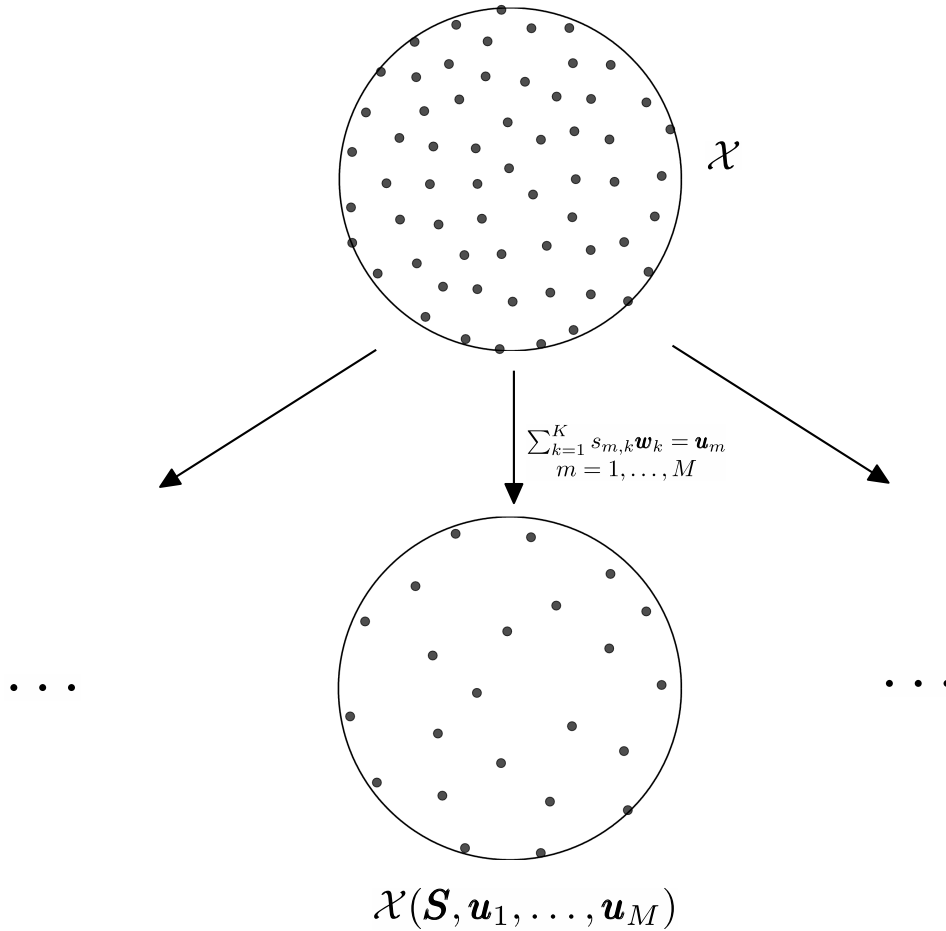
Fig. 3. Each receiver $(\boldsymbol{S}, \sigma^2)$ of the multicast channel uses its own side information $(\boldsymbol{S}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_M)$ to expurgate the channel code $\mathcal{X}$ and obtain a subcode $\mathcal{X}(\boldsymbol{S}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_M)$. Note that the resulting subcodes can be different across the receivers. In order to achieve the capacity of the multicast channel, we require that each of these expurgated codes be good for channel coding over the point-to-point AWGN channel.

$m = 1, \dots, M$, where $M = \text{rank}(\boldsymbol{S})$ and $\boldsymbol{u}_m \in \mathbb{F}_p^\ell$. This side information allows the receiver to conclude that the transmitted codeword must belong to the following subcode $\mathcal{X}(\boldsymbol{S}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_M)$ of $\mathcal{X}$,

$$\Big\{ \rho(\boldsymbol{w}_1, \dots, \boldsymbol{w}_K) \ \Big| \ \boldsymbol{w}_1, \dots, \boldsymbol{w}_K \in \mathbb{F}_p^\ell,$$
$$\sum_{k=1}^K s_{m,k} \boldsymbol{w}_k = \boldsymbol{u}_m \text{ for } m = 1, \dots, M \Big\}. \quad (3)$$

The optimal decoder at $(\boldsymbol{S}, \sigma^2)$ decodes the channel output vector to the nearest codeword $\hat{\boldsymbol{x}}$ of this subcode, and the error probability at this receiver is the probability that the estimated message tuple $\rho^{-1}(\hat{\boldsymbol{x}})$ is not equal to the transmit message $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_K)$. In order to achieve the optimal performance at a given receiver $(\boldsymbol{S}, \sigma^2)$, we thus require that the expurgated code $\mathcal{X}(\boldsymbol{S}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_M)$ be a good channel code for the point-to-point AWGN channel. In the multicast channel that consists of multiple receivers, the side information matrix $\boldsymbol{S}$ can vary from one receiver to the next, and hence, the expurgated codes can be different at each receiver, see Fig. 3. Hence, a capacity-achieving channel code $\mathcal{X}$ is such that the resulting expurgated code at every receiver is a good channel code for the AWGN channel.

Problem Statement

Problem Setup: Consider a common message Gaussian broadcast channel with single transmitter and N receivers. The transmitter desires to multicast K independent messages from a prime field $\mathbb{F}_p$ subject to the unit power constraint (2) on the transmit codeword. Each of the N receivers $(\boldsymbol{S}_1, \sigma_1^2), \dots, (\boldsymbol{S}_N, \sigma_N^2)$ has coded side information corresponding to the side information matrix $\boldsymbol{S}_i \in \mathbb{F}_p^{M_i \times K}$, $i = 1, \dots, N$, and experiences an additive white Gaussian noise of variance σ_i^2, $i = 1, \dots, N$. Without loss of generality, we assume that each of the side information matrices $\boldsymbol{S}_i$ has linearly independent rows, i.e., $\text{rank}(\boldsymbol{S}_i) = M_i$. Using the information-theoretic arguments of [17], which is based on the average performance of an ensemble of randomly generated codebooks, it can be shown that the (symmetric) capacity of this multicast channel is

$$C = \min_{i \in \{1, \dots, N\}} \frac{1}{(K - M_i)} \cdot \frac{1}{2} \log_2 \left(1 + \frac{1}{\sigma_i^2} \right). \quad (4)$$

The proof of this result is similar to the proof of [17, Th. 6] which considers a *discrete memoryless* common message broadcast channel where the side information at each receiver

is, in general, a random variable jointly distributed with the source messages $(w_1, \ldots, w_K)$. A sketch of the proof that C is the capacity for the Gaussian multicast channel with coded side information at the receivers is given in the appendix.

Problem Statement: Let $\epsilon, \zeta > 0$ be fixed positive real numbers and let $R \leq C - \epsilon$. We seek to determine whether there exists a lattice code for the multicast channel with coded side information at the receivers that transmits each of the K messages with rate at least $(R - \epsilon)$ such that the probability of decoding error at each of the N receivers is at the most ζ.

In this paper we answer the above stated problem in the affirmative under the assumption that the prime field $\mathbb{F}_p$ is sufficiently large. In particular, we prove the existence of a lattice code with the said properties when the prime p satisfies the inequality $p \geq \max \left\{ 2^{2KR}, (2^{\frac{\epsilon}{4}} - 1)^{-1} 2^{-R} \right\}$. Unlike the capacity (4) which holds for any value of p, our result on the optimality of lattice codes requires that p vary with the tolerance ϵ. The larger the gap to capacity ϵ, the smaller is the size requirement on the prime field $\mathbb{F}_p$.

B. Lattice Preliminaries

We now briefly recall the necessary properties of lattices and lattice codes, and establish our notation and terminology. The material presented in this section consists of standard ingredients used in the literature, and is mainly based on [5], [8], [42], and [43].

1) Lattices and Lattice Codes: Throughout this manuscript we consider n-dimensional lattices Λ with full-rank generator matrices. The closest vector lattice quantizer corresponding to Λ is denoted by the function $Q_\Lambda : \mathbb{R}^n \to \Lambda$, and the volume of its (fundamental) Voronoi region $\mathcal{V}(\Lambda) = Q_\Lambda^{-1}(\mathbf{0})$ is denoted by $\text{Vol}(\Lambda)$. For any $\boldsymbol{\lambda} \in \Lambda$, $\boldsymbol{\lambda} + \mathcal{V}(\Lambda)$ is the set of all points in $\mathbb{R}^n$ that are mapped to $\boldsymbol{\lambda}$ under Q_Λ, and it has the same volume $\text{Vol}(\Lambda)$ as $\mathcal{V}(\Lambda)$. For any two distinct lattice points $\boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_2$, the sets $\boldsymbol{\lambda}_1 + \mathcal{V}(\Lambda)$ and $\boldsymbol{\lambda}_2 + \mathcal{V}(\Lambda)$ are disjoint. The *modulo-Λ* operation, defined as $[\boldsymbol{x}] \bmod \Lambda = \boldsymbol{x} - Q_\Lambda(\boldsymbol{x})$, satisfies the following properties for all $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$

$$[\boldsymbol{x}] \bmod \Lambda \in \mathcal{V}(\Lambda),$$

$$[\boldsymbol{x}_1 + \boldsymbol{x}_2] \bmod \Lambda = \left[[\boldsymbol{x}_1] \bmod \Lambda + \boldsymbol{x}_2 \right] \bmod \Lambda, \text{ and} \quad (5)$$

$$[\boldsymbol{x}] \bmod \Lambda = \mathbf{0} \text{ if and only if } \boldsymbol{x} \in \Lambda. \quad (6)$$

We will denote the n-dimensional ball of radius r with center $\boldsymbol{s} \in \mathbb{R}^n$ as $\mathcal{B}(\boldsymbol{s}, r)$, i.e. $\mathcal{B}(\boldsymbol{s}, r) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{s}\| \leq r\}$, and the volume of a unit-radius ball in n dimensions by V_n. It follows that the volume of $\mathcal{B}(\boldsymbol{s}, r)$ equals $V_n r^n$. The *covering radius* of the lattice Λ is denoted by $r_{\text{cov}}(\Lambda)$ and the *effective radius* of Λ by $r_{\text{eff}}(\Lambda)$. We recall that $r_{\text{eff}}(\Lambda) \leq r_{\text{cov}}(\Lambda)$ and

$$r_{\text{eff}}(\Lambda) = \left(\frac{\text{Vol}(\Lambda)}{V_n} \right)^{\frac{1}{n}}. \quad (7)$$

Rogers [44] showed that for every dimension n there exists a lattice Λ such that

$$\frac{r_{\text{cov}}(\Lambda)}{r_{\text{eff}}(\Lambda)} \leq \left(c \, n \, (\ln n)^{\frac{1}{2} \log 2\pi e} \right)^{\frac{1}{n}}, \quad (8)$$

where c is a constant. Note that the right hand side of the above inequality converges to 1 as $n \to \infty$. A sequence of

lattices of increasing dimension n is said to be *Rogers-good* if $\frac{r_{\text{cov}}}{r_{\text{eff}}} \to 1$. Rogers' result (8) shows that such a sequence exists (see also [45]).

Let $\Lambda_c \subset \Lambda$ be a pair of nested lattices and $\boldsymbol{d} \in \mathcal{V}(\Lambda_c)$ be a fixed vector. A *(nested) lattice code* or a *Voronoi code* $(\Lambda - \boldsymbol{d})/\Lambda_c$ is the set $(\Lambda - \boldsymbol{d}) \bmod \Lambda_c$ obtained by applying the $\bmod \Lambda_c$ operation on the points of the lattice translate $\Lambda - \boldsymbol{d}$. The code consists of all the points in $\Lambda - \boldsymbol{d}$ that lie within the Voronoi region of Λ_c, i.e., $(\Lambda - \boldsymbol{d}) \cap \mathcal{V}(\Lambda_c)$. The lattice Λ_c is called the *coarse lattice* or the *shaping lattice*, Λ is called the *fine lattice* or the *coding lattice*, and $\boldsymbol{d}$ is the *dither* vector. The cardinality of this code is $|(\Lambda - \boldsymbol{d})/\Lambda_c| = |\Lambda/\Lambda_c| = \frac{\text{Vol}(\Lambda_c)}{\text{Vol}(\Lambda)}$, and every codeword point $\boldsymbol{x} \in (\Lambda - \boldsymbol{d})/\Lambda_c$ satisfies $\|\boldsymbol{x}\| \leq r_{\text{cov}}(\Lambda_c)$. Note that Λ/Λ_c is a lattice code with zero dither.

2) Lattice Codes from Linear Codes over a Finite Field: In this subsection we briefly describe the method proposed in [15] to construct a pair $\Lambda_c \subset \Lambda$ of nested lattices, and recall its relevant properties. This construction uses a coarse lattice Λ_c and a linear code $\mathscr{C}$ to generate a fine lattice Λ such that $|\Lambda/\Lambda_c| = |\mathscr{C}|$.

Let $g(\cdot)$ denote the natural map that embeds $\mathbb{F}_p = \{0, 1, \ldots, p - 1\}$ into $\mathbb{Z}$. When applied to vectors, $g(\cdot)$ acts independently on each component of a vector. Let $\mathscr{C} \subset \mathbb{F}_p^n$ be a linear code of rank L, $1 \leq L \leq n$,

$$\mathscr{C} = \left\{ \boldsymbol{G}\boldsymbol{w} \mid \boldsymbol{w} \in \mathbb{F}_p^L \right\},$$

where $\boldsymbol{G}$ is the $n \times L$ generator matrix with full column rank, and $\boldsymbol{w}$ is the message encoded to $\mathscr{C}$. The set $g(\mathscr{C}) + p\mathbb{Z}^n$ obtained by tiling copies of $g(\mathscr{C})$ at every vector of $p\mathbb{Z}^n$ is a lattice in $\mathbb{R}^n$ and is known as the *Construction A* lattice of the linear code $\mathscr{C}$ [5]. Note that the number of points in $g(\mathscr{C}) + p\mathbb{Z}^n$ contained in the Voronoi region of the lattice $p\mathbb{Z}^n$ is $|\mathscr{C}| = p^L$. We obtain Λ by scaling down the Construction A lattice by p^{-1} and transforming it by the generator matrix $\boldsymbol{B}_c$ of Λ_c

$$\Lambda = \boldsymbol{B}_c p^{-1} \left(g(\mathscr{C}) + p\mathbb{Z}^n \right) = \boldsymbol{B}_c p^{-1} g(\mathscr{C}) + \boldsymbol{B}_c \mathbb{Z}^n$$
$$= \boldsymbol{B}_c p^{-1} g(\mathscr{C}) + \Lambda_c.$$

Since $\mathscr{C}$ contains the all zero codeword, it follows that $\Lambda \supset \boldsymbol{B}_c p^{-1} g(\mathbf{0}) + \Lambda_c = \Lambda_c$. We observe that applying the transformation $\boldsymbol{B}_c p^{-1}$ to the lattice $p\mathbb{Z}^n$ (instead of the lattice $g(\mathscr{C}) + p\mathbb{Z}^n$) generates Λ_c (instead of Λ). Hence, Λ/Λ_c has the same algebraic structure as that of $(g(\mathscr{C}) + p\mathbb{Z}^n)/p\mathbb{Z}^n$, which in turn, is equivalent to the linear code $\mathscr{C}$. In particular,

$$|\Lambda/\Lambda_c| = |\mathscr{C}| = p^L. \quad (9)$$

The following lemma provides an explicit bijection between the message vectors $\boldsymbol{w} \in \mathbb{F}_p^L$ encoded by $\mathscr{C}$ and the points in the lattice code Λ/Λ_c. This result, which is originally from [15, Lemma 5], is proved below for completeness.

Lemma 1: The map $\boldsymbol{w} \to \left[\boldsymbol{B}_c \, p^{-1} g \left(\boldsymbol{G}\boldsymbol{w} \right) \right] \bmod \Lambda_c$ *is a bijection between* $\mathbb{F}_p^L$ *and* Λ/Λ_c.

Proof: From (9), we know that $|\Lambda/\Lambda_c| = |\mathbb{F}_p^L| = p^L$. Hence, it only remains to show that no two distinct messages $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$ are mapped to the same point in Λ/Λ_c.

Assuming the contrary, we have $\left[\boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w}_A)\right]$ mod $\Lambda_c = \left[\boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w}_B)\right]$ mod Λ_c. Using (5) and (6), we obtain

$$\boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w}_A) - \boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w}_B) \in \Lambda_c.$$

Multiplying both sides by $p\boldsymbol{B}_c^{-1}$, we obtain $g(\boldsymbol{G}\boldsymbol{w}_A) - g(\boldsymbol{G}\boldsymbol{w}_B) \in p\mathbb{Z}^n$. Reducing this result modulo-p, we have $\boldsymbol{G}\boldsymbol{w}_A - \boldsymbol{G}\boldsymbol{w}_B = \boldsymbol{0}$ over $\mathbb{F}_p$. Since this implies $\boldsymbol{G}(\boldsymbol{w}_A - \boldsymbol{w}_B) = \boldsymbol{0}$ over $\mathbb{F}_p$ while $\boldsymbol{w}_A - \boldsymbol{w}_B \neq \boldsymbol{0}$ and $\boldsymbol{G}$ has full column rank, we have arrived at a contradiction. ∎

In order to prove capacity achievability, we will rely on random coding arguments to show the existence of a good choice of $\boldsymbol{G}$. As in [15], we will assume that $\boldsymbol{G}$ is a random matrix chosen with uniform probability distribution on $\mathbb{F}_p^{n \times L}$. The following result is useful in upper bounding the decoding error probability over the ensemble of random codes.

Lemma 2 [15], [41], [45]: Let $\boldsymbol{w} \in \mathbb{F}_p^L \setminus \{\boldsymbol{0}\}$ be a given non-zero vector, and let $\boldsymbol{G}$ be uniformly distributed in $\mathbb{F}_p^{n \times L}$. Then $\left[\boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w})\right]$ mod Λ_c is uniformly distributed over $\left(p^{-1}\Lambda_c\right) \cap \mathcal{V}(\Lambda_c)$, i.e., over the lattice code $p^{-1}\Lambda_c / \Lambda_c$.

III. LATTICE CODES FOR THE COMMON MESSAGE GAUSSIAN BROADCAST CHANNEL WITH CODED SIDE INFORMATION

We will assume that the number of messages K and a design rate R are given, and show that there exist good lattice codes of sufficiently large dimension n that encode K messages over an appropriately chosen prime field $\mathbb{F}_p$ at rates close to R b/dim. In order to rigorously state the main result, we consider a fixed non-zero tolerance $\epsilon > 0$ that determines the gap to capacity.

Theorem 1 (Main Theorem): Let the number of messages K, design rate R and tolerance $\epsilon > 0$ be given. For every sufficiently large prime integer p, there exists a sequence of lattice codes of increasing dimension n that encode K message vectors over $\mathbb{F}_p$ such that the rate of transmission of each message is at least $(R - \epsilon)$ b/dim and the probability of error at a receiver $(\boldsymbol{S}, \sigma^2)$ decays exponentially in n if

$$\frac{1}{2} \log \left(1 + \frac{1}{\sigma^2}\right) \geq (R + \epsilon)(K - \mathrm{rank}(\boldsymbol{S})). \tag{10}$$

To prove Theorem 1, we utilize the lattice code ensemble introduced in [15]; see Section II-B2 of this paper. A Rogers' good lattice is chosen as the coarse lattice Λ_c. The fine lattice Λ is obtained from the generator matrix $\boldsymbol{B}_c$ of the coarse lattice Λ_c and a linear code $\mathscr{C}$ over a large enough prime field $\mathbb{F}_p$ using the construction described in Section II-B2.

The multicast channel considered in Theorem 1 reduces to the traditional single-user AWGN channel if the number of messages $K = 1$, and the multicast channel consists of one receiver with an empty side information matrix $\boldsymbol{S}$, i.e., $\mathrm{rank}(\boldsymbol{S}) = 0$. Hence, Theorem 1 provides an alternative proof of the existence of lattice codes that achieve the capacity of the single-user AWGN channel, and we have the following corollary.

Corollary 1: Consider a single user AWGN channel where the input is subject to unit power constraint and the noise variance at the receiver is σ^2. Let $\epsilon > 0$ be any constant and let

$$R \leq \frac{1}{2} \log \left(1 + \frac{1}{\sigma^2}\right) - \epsilon.$$

For every sufficiently large prime integer p, there exists a sequence of lattice codes of increasing dimension n constructed from linear codes over $\mathbb{F}_p$ (as described in Section II-B2) such that the rate of each lattice code is at least $R - \epsilon$ and the probability of decoding error at the receiver decays exponentially in n.

The relation of Corollary 1 to existing results on the optimality of Construction A based lattice codes in single-user AWGN channel is described in detail in Section IV-D2.

In the rest of this section we describe the construction of random lattice codes, and the encoding and decoding operations used to prove Theorem 1. We provide the proof of the Theorem 1 in Section IV.

A. Random Lattice Code Ensemble

1) Prime p: Given the design rate R, number of messages K and tolerance $\epsilon > 0$, we require p to satisfy the constraint

$$p \geq \max \left\{2^{2KR}, (2^{\frac{\epsilon}{4}} - 1)^{-1} 2^{-R}\right\}. \tag{11}$$

The coding schemes of this paper are based on Construction A lattices which are obtained by lifting linear codes over $\mathbb{F}_p$ to the Euclidean space $\mathbb{R}^n$. The generator matrices of these $\mathbb{F}_p$-linear codes are constructed randomly, and the first constraint in (11), viz. $p \geq 2^{2KR}$, will allow us to show that these randomly constructed generator matrices are full-ranked with probability close to 1.

The proof of Theorem 1 given in Section IV involves the derivation of an upper bound on the probability of decoding error averaged over an ensemble of lattice codes derived from Construction A. We will use the inequality $p \geq (2^{\frac{\epsilon}{4}} - 1)^{-1} 2^{-R}$ from (11) to show that this upper bound is exponentially small in dimension n. Note that this inequality implies

$$p \geq (2^{\frac{\epsilon}{4}} - 1)^{-1} 2^{-(R+\epsilon)(K-M)}$$

for any integer M satisfying $0 \leq M \leq K - 1$. Rearranging the terms in the above inequality we obtain

$$\frac{1}{p \, 2^{(R+\epsilon)(K-M)}} + 1 \leq 2^{\frac{\epsilon}{4}}. \tag{12}$$

2) Message Length ℓ: Once p is fixed, we choose ℓ as the largest integer that satisfies

$$\frac{\ell}{n} \log p \leq R. \tag{13}$$

The left-hand side in the above inequality is the actual rate at which the lattice code encodes each message, while R is the design rate. The difference between the two is at the most

$$\frac{\ell + 1}{n} \log p - \frac{\ell}{n} \log p = \frac{\log p}{n}$$

which converges to 0 as $n \to \infty$. It follows that the code rate $\frac{\ell}{n} \log_2 p$ tends to the design rate R as $n \to \infty$, and hence, $\frac{\ell}{n} \log_2 p \geq R - \epsilon$ for all sufficiently large n.
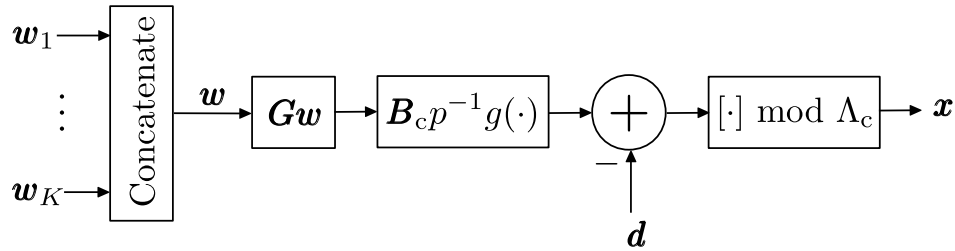
Fig. 4.   The encoding operation at the transmitter that maps the messages $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ to a point in $(\Lambda - \boldsymbol{d})/\Lambda_c$.

3) Coarse Lattice Λ_c: From (8) in Section II-B1, we know that for a given $\epsilon > 0$ and for all sufficiently large n, there exists an n-dimensional lattice Λ_c such that

$$\frac{r_{\mathrm{cov}}(\Lambda_c)}{r_{\mathrm{eff}}(\Lambda_c)} \leq 2^{\frac{\epsilon}{4}}.$$

We will choose such a Rogers-good lattice as Λ_c, and scale it so that

$$r_{\mathrm{cov}}(\Lambda_c) = \sqrt{n}.$$

It follows that $r_{\mathrm{eff}}(\Lambda_c) \geq 2^{-\frac{\epsilon}{4}} r_{\mathrm{cov}}(\Lambda_c) = 2^{-\frac{\epsilon}{4}}\sqrt{n}$. Using the definition of the effective radius (7), we arrive at the following lower bound on the volume of the Voronoi region of Λ_c

$$\mathrm{Vol}(\Lambda_c) = V_n\, r_{\mathrm{eff}}^n(\Lambda_c) \geq V_n\, n^{\frac{n}{2}}\, 2^{-\frac{n\epsilon}{4}}. \tag{14}$$

4) Fine Lattice Λ: The fine lattice is obtained by the construction of [15] described in Section II-B2. The length of the linear code $\mathscr{C}$ is n, and its rank $L = K\ell$ is the number of message symbols to be encoded by the lattice code. Note that this requires that $K\ell < n$ be true. Using (13) and the property $p \geq 2^{2KR}$, we have

$$K\ell \leq \frac{nKR}{\log p} \leq \frac{nKR}{2KR} = \frac{n}{2} \tag{15}$$

which ensures that $K\ell < n$. If $\boldsymbol{G} \in \mathbb{F}_p^{n \times K\ell}$ is the generator matrix of $\mathscr{C}$, then $\Lambda = \boldsymbol{B}_c p^{-1} g(\mathscr{C}) + \Lambda_c$. We will choose $\boldsymbol{G}$ uniformly random over the set of all $n \times K\ell$ matrices of $\mathbb{F}_p$, resulting in a random ensemble of fine lattices Λ.

5) Dither Vector $\boldsymbol{d}$: We will rely on random coding arguments to prove the existence of a translate $\boldsymbol{d}$ such that the code $(\Lambda - \boldsymbol{d})/\Lambda_c$ performs close to capacity. We will assume that $\boldsymbol{d}$ is distributed uniformly in $\mathcal{V}(\Lambda_c)$ and is chosen independently of $\boldsymbol{G}$. This random dither $\boldsymbol{d}$ is usually viewed as a common randomness available at the transmitter and the receivers [8]. Note that $\|\boldsymbol{d}\| \leq r_{\mathrm{cov}}(\Lambda_c) = \sqrt{n}$.

B. Encoding

We will now describe the encoding operation ρ at the transmitter that maps the message vectors $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) \in \mathbb{F}_p^\ell \times \cdots \times \mathbb{F}_p^\ell$ to a codeword $\boldsymbol{x} \in (\Lambda - \boldsymbol{d})/\Lambda_c$. The encoder first concatenates the K messages into the vector $\boldsymbol{w} = (\boldsymbol{w}_1^\mathsf{T}, \cdots, \boldsymbol{w}_K^\mathsf{T})^\mathsf{T}$, encodes $\boldsymbol{w}$ to a codeword in the linear code $\mathscr{C}$, and maps it to a point $\boldsymbol{t} \in \mathbb{R}^n$ using Construction A as follows

$$\boldsymbol{t} = \left[\boldsymbol{B}_c\, p^{-1} g\,(\boldsymbol{G}\boldsymbol{w}) \right] \bmod \Lambda_c. \tag{16}$$

From the discussion in Section III-A, we know that $\boldsymbol{B}_c p^{-1} g(\boldsymbol{G}\boldsymbol{w}) \in \Lambda$, and hence, $\boldsymbol{t} \in \Lambda/\Lambda_c$. Finally, the transmit codeword $\boldsymbol{x}$ is generated by dithering $\boldsymbol{t}$,

$$\boldsymbol{x} = [\boldsymbol{t} - \boldsymbol{d}] \bmod \Lambda_c = \left[\boldsymbol{B}_c\, p^{-1} g\,(\boldsymbol{G}\boldsymbol{w}) - \boldsymbol{d} \right] \bmod \Lambda_c. \tag{17}$$

This sequence of operations is illustrated in Fig. 4. Note that since $r_{\mathrm{cov}}(\Lambda_c) = \sqrt{n}$, each codeword $\boldsymbol{x}$ satisfies $\|\boldsymbol{x}\| \leq r_{\mathrm{cov}}(\Lambda_c) = \sqrt{n}$, and hence, the power constraint $\|\boldsymbol{x}\|^2/n \leq 1$. It is straightforward to show that the dithering operation (17) is a one-to-one correspondence between $\boldsymbol{t} \in \Lambda/\Lambda_c$ and $\boldsymbol{x} \in (\Lambda - \boldsymbol{d})/\Lambda_c$. Further, from Lemma 1 we know that (16) is a bijection between the message space $\mathbb{F}_p^{K\ell}$ and the undithered codewords Λ/Λ_c if $\boldsymbol{G}$ is full rank. Hence, to ensure that no two messages are mapped to the same codeword, we only require that the random matrix $\boldsymbol{G}$ be full rank. It can be shown that (see [45])

$$\mathrm{P}\,(\mathrm{rank}(\boldsymbol{G}) < K\ell) \leq p^{-(n-K\ell)}.$$

We will only require a relaxation based on the above inequality. From (15), we have $K\ell \leq \frac{n}{2}$. Similarly, since p is a prime integer, we have $p \geq 2$, and hence,

$$\mathrm{P}\,(\mathrm{rank}(\boldsymbol{G}) < K\ell) \leq 2^{-\left(n-\frac{n}{2}\right)} = 2^{-\frac{n}{2}}. \tag{18}$$

C. Decoding

The receiver employs a two stage decoder: in the first stage the receiver identifies the subcode of $(\Lambda - \boldsymbol{d})/\Lambda_c$ corresponding to the available side information, and in the second stage it decodes the channel output to a point in this subcode.

1) Using Side Information to Expurgate Codewords: The side information at $(\boldsymbol{S}, \sigma^2)$ over a block of ℓ realizations of the K messages is of the form

$$\sum_{k=1}^{K} s_{m,k}\boldsymbol{w}_k = \boldsymbol{u}_m, \quad m = 1, \ldots, M. \tag{19}$$

The receiver desires to identify the set of all possible values of the message vector $\boldsymbol{w} = (\boldsymbol{w}_1^\mathsf{T}, \cdots, \boldsymbol{w}_K^\mathsf{T})^\mathsf{T}$ that satisfy (19). Using the notation $\boldsymbol{u} = (\boldsymbol{u}_1^\mathsf{T}, \cdots, \boldsymbol{u}_M^\mathsf{T})^\mathsf{T} \in \mathbb{F}_p^{M\ell}$, the side information (19) can be rewritten compactly in terms of $\boldsymbol{w}$ and $\boldsymbol{u}$ as

$$(\boldsymbol{S} \otimes \boldsymbol{I}_\ell)\,\boldsymbol{w} = \boldsymbol{u}, \tag{20}$$

where $\otimes$ denotes the Kronecker product of matrices and $\boldsymbol{I}_\ell$ is the $\ell \times \ell$ identity matrix over $\mathbb{F}_p$. Observe that (20) is an under-determined system of linear equations, and the set of solutions

is a coset of the null space of $S \otimes I_\ell$. Let $A_S \in \mathbb{F}_p^{K\ell \times (K-M)\ell}$ be a rank $(K-M)\ell$ matrix such that $(S \otimes I_\ell)A_S = 0$, i.e., the columns of A_S form a basis of the null space of $S \otimes I_\ell$. Then the set of all solutions to (20) is

$$v + \left\{ A_S \tilde{w} \mid \tilde{w} \in \mathbb{F}_p^{(K-M)\ell} \right\}, \tag{21}$$

where v is the coset leader. From (16), we conclude that the undithered codeword must be of the form

$$t = \left[B_c p^{-1} g \left(Gv + GA_S \tilde{w} \right) \right] \bmod \Lambda_c, \quad \tilde{w} \in \mathbb{F}_p^{(K-M)\ell}. \tag{22}$$

We will now use the property of $g(\cdot)$ that for any $a, b \in \mathbb{F}_p^n$,

$$g(a + b) = g(a) + g(b) \bmod p.$$

Therefore, $g(Gv + GA_S \tilde{w}) = g(Gv) + g(GA_S \tilde{w}) + pc$ for some $c \in \mathbb{Z}^n$. Using this in (22), we obtain

$$
\begin{aligned}
t &= \left[B_c p^{-1} g(Gv) + B_c p^{-1} g(GA_S \tilde{w}) + B_c c \right] \bmod \Lambda_c \\
&= \left[B_c p^{-1} g(Gv) + \left[B_c p^{-1} g(GA_S \tilde{w}) \right] \bmod \Lambda_c \right. \\
&\quad \left. + [B_c c] \bmod \Lambda_c \right] \bmod \Lambda_c \\
&= \left[B_c p^{-1} g(Gv) + \left[B_c p^{-1} g(GA_S \tilde{w}) \right] \bmod \Lambda_c \right] \bmod \Lambda_c,
\end{aligned}
\tag{23}
$$

where we have used (5), (6) and the fact that $B_c c \in \Lambda_c$. Since the receiver knows v, the component of t unavailable from the side information is

$$\tilde{t} = \left[B_c p^{-1} g \left(GA_S \tilde{w} \right) \right] \bmod \Lambda_c. \tag{24}$$

Let $\mathscr{C}_S \subset \mathbb{F}_p^n$ be the subcode of $\mathscr{C}$ with generator matrix GA_S, and Λ_S be the lattice obtained by applying Construction A to $\mathscr{C}_S$ and transforming it by $B_c p^{-1}$, i.e.,

$$\Lambda_S = B_c p^{-1} g(\mathscr{C}_S) + \Lambda_c.$$

Using GA_S instead of G in Lemma 1, we see that $\tilde{t} \in \Lambda_S / \Lambda_c$ and that (24) is a one-to-one correspondence between $\tilde{w} \in \mathbb{F}_p^{(K-M)\ell}$ and $\tilde{t} \in \Lambda_S / \Lambda_c$ as long as GA_S is full rank. Together with (17), (23), and (24), we conclude that the transmit vector x belongs to the following lattice subcode of $(\Lambda - d)/\Lambda_c$,

$$
\begin{aligned}
&\left(\Lambda_S + B_c p^{-1} g(Gv) - d \right) / \Lambda_c \\
&= \left\{ \left[\tilde{t} + B_c p^{-1} g(Gv) - d \right] \bmod \Lambda_c \mid \tilde{t} \in \Lambda_S / \Lambda_c \right\}.
\end{aligned}
\tag{25}
$$

The decoding problem at the second stage is to estimate $\tilde{t}$, or equivalently $\tilde{w}$, from the channel output.

2) MMSE Scaling and Lattice Decoding: Let the channel output at the receiver (S, σ^2) be $y = x + n$, where n is a Gaussian vector with zero mean and variance σ^2 per dimension. The received vector is scaled by the coefficient α, resulting in

$$\alpha y = \alpha x + \alpha n = x + \alpha n - (1 - \alpha)x. \tag{26}$$

This MMSE pre-processing improves the effective signal-to-noise ratio of the system beyond the channel signal-to-noise

ratio $\frac{1}{\sigma^2}$ and allows the lattice decoder to perform close to capacity [8], [46]. Let

$$z = \alpha n - (1 - \alpha)x$$

be the effective noise term in (26). Using the facts that x and n are independent, $\|x\| \le \sqrt{n}$, and n has zero mean, we have

$$
\begin{aligned}
\mathbb{E}\|z\|^2 &= (1-\alpha)^2 \, \mathbb{E}\|x\|^2 + \alpha^2 \mathbb{E}\|n\|^2 - 2\alpha(1-\alpha)\,\mathbb{E}\,x^\mathsf{T} n \\
&\le (1-\alpha)^2 \, n + \alpha^2 \sigma^2 \, n,
\end{aligned}
$$

where $\mathbb{E}$ is the expectation operator. The choice of $\alpha = 1/(1 + \sigma^2)$ minimizes this upper bound and yields $\mathbb{E}\|z\|^2 \le n \, \sigma^2/(1 + \sigma^2)$, which is less than the Gaussian noise power $\mathbb{E}\|n\|^2 = n\sigma^2$. In the rest of the paper we will assume that $\alpha = 1/(1 + \sigma^2)$ and use the notation

$$\sigma_z^2 = \frac{\sigma^2}{1 + \sigma^2}. \tag{27}$$

The lower bound (10) on signal-to-noise ratio can be rewritten in terms of σ_z^2 as

$$\sigma_z^2 \le 2^{-2(R+\epsilon)(K-M)}. \tag{28}$$

From (25), we know that $x = \tilde{t} + B_c p^{-1} g(Gv) - d + \lambda_c$ for some $\lambda_c \in \Lambda_c$. After MMSE scaling, the decoder removes the contributions of the dither d and the offset $B_c p^{-1} g(Gv)$ from αy to obtain

$$y' = \alpha y - B_c p^{-1} g(Gv) + d = \tilde{t} + \lambda_c + z.$$

The decoder proceeds by quantizing y' to the lattice Λ_S and reducing the result modulo Λ_c. If the noise z is sufficiently 'small', then this sequence of operations will yield

$$
\begin{aligned}
\left[Q_{\Lambda_S}(y') \right] \bmod \Lambda_c &= \left[Q_{\Lambda_S}(\tilde{t} + \lambda_c + z) \right] \bmod \Lambda_c \\
&= \left[\tilde{t} + \lambda_c \right] \bmod \Lambda_c = \tilde{t}.
\end{aligned}
\tag{29}
$$

Given $\tilde{t}$, the receiver uses (23) to obtain the undithered codeword t, and hence the message vector $(w_1^\mathsf{T}, \ldots, w_K^\mathsf{T})^\mathsf{T}$, as $t = \left[B_c p^{-1} g(Gv) + \tilde{t} \right] \bmod \Lambda_c$. To conclude, the decoder obtains the estimate $\hat{t}$ of the undithered codeword t from the received vector y as

$$
\begin{aligned}
\hat{t} &= \left[\left[Q_{\Lambda_S}(y') \right] \bmod \Lambda_c + B_c p^{-1} g(Gv) \right] \bmod \Lambda_c \\
&= \left[Q_{\Lambda_S}(y') + B_c p^{-1} g(Gv) \right] \bmod \Lambda_c \\
&= \left[Q_{\Lambda_S}\left(\alpha y - B_c p^{-1} g(Gv) + d \right) + B_c p^{-1} g(Gv) \right] \bmod \Lambda_c
\end{aligned}
$$

which shows that the mod Λ_c operation arising from (29) can be ignored. The steps involved in the decoding operation are illustrated in Fig. 5.

Note that the effective information vector $\tilde{w}$ is not encoded in the point $\tilde{t} \in \Lambda_S$, but is encoded in the coset $\tilde{t} + \Lambda_c$. The error event for this decoder is $Q_{\Lambda_S}(y') \notin \tilde{t} + \Lambda_c$, i.e., $Q_{\Lambda_S}(\tilde{t} + \lambda_c + z) \notin \tilde{t} + \Lambda_c$, which is equivalent to the event $Q_{\Lambda_S}(z) \notin \Lambda_c$. Hence, a decoding error occurs if and only if z is closer to a point in $\Lambda_S \backslash \Lambda_c$ than any vector in the coarse lattice Λ_c, i.e., if and only if

$$\mathcal{E}: Q_{\Lambda_S}(z) \in \Lambda_S \backslash \Lambda_c. \tag{30}$$
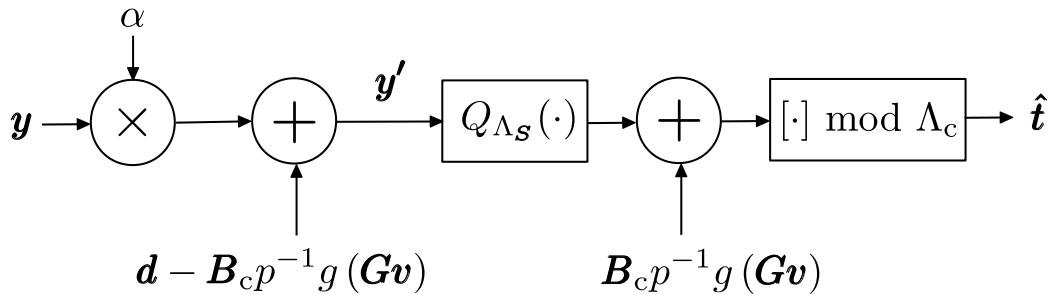
Fig. 5. The decoding operation at the receiver $(\boldsymbol{S}, \sigma^2)$. The vector $\boldsymbol{v}$ and the lattice $\Lambda_{\boldsymbol{S}}$ are determined using the side information available at the receiver.

IV. PROOF OF MAIN THEOREM

In this section we first state and prove two technical lemmas (Section IV-A), use these lemmas to show that the error probability at a given fixed receiver $(\boldsymbol{S}, \sigma^2)$ is small (Section IV-B), and then complete the proof of the main theorem by showing that the error probability at every receiver of the multicast channel is simultaneously small (Section IV-C). Finally, we state some important corollaries of the main theorem (Section IV-D).

A. Technical Lemmas

The first result, which is a direct generalization of [9, Lemma 1] and [10, Lemma 2.3], gives an upper bound on the number of lattice points lying inside a ball.

Lemma 3: For any $\boldsymbol{s} \in \mathbb{R}^n$, $r > 0$ and any n-dimensional lattice Λ_c,

$$|\Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)| \leq \frac{V_n}{\mathrm{Vol}(\Lambda_c)} (r_{\mathrm{cov}}(\Lambda_c) + r)^n,$$

where V_n is the volume of a unit ball in $\mathbb{R}^n$.

Proof: Let $\mathcal{R} = (\Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)) + \mathcal{V}(\Lambda_c)$ be the set of all points in $\mathbb{R}^n$ that are mapped to one of the points in $\Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)$ by the lattice quantizer Q_{Λ_c}. Since $\mathcal{R}$ is a union of the pairwise disjoint sets $\boldsymbol{\lambda} + \mathcal{V}(\Lambda_c)$, $\boldsymbol{\lambda} \in \Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)$, and since each of these sets has volume $\mathrm{Vol}(\Lambda_c)$, we have

$$\mathrm{Vol}(\mathcal{R}) = \mathrm{Vol}(\Lambda_c) \, |\Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)|. \tag{31}$$

Using the fact that $\mathcal{V}(\Lambda_c) \subset \mathcal{B}(\boldsymbol{0}, r_{\mathrm{cov}}(\Lambda_c))$, we have

$$\mathcal{R} = (\Lambda_c \cap \mathcal{B}(\boldsymbol{s}, r)) + \mathcal{V}(\Lambda_c) \subset \mathcal{B}(\boldsymbol{s}, r) + \mathcal{V}(\Lambda_c)$$
$$\subset \mathcal{B}(\boldsymbol{s}, r) + \mathcal{B}(\boldsymbol{0}, r_{\mathrm{cov}}(\Lambda_c)) \subset \mathcal{B}(\boldsymbol{s}, r + r_{\mathrm{cov}}(\Lambda_c)),$$

where the last step follows from triangle inequality. Consequently, we have an upper bound on the volume of $\mathcal{R}$,

$$\mathrm{Vol}(\mathcal{R}) \leq \mathrm{Vol}\left(\mathcal{B}(\boldsymbol{s}, r + r_{\mathrm{cov}}(\Lambda_c))\right) = V_n (r + r_{\mathrm{cov}}(\Lambda_c))^n.$$

Using this result with (31) proves the lemma. ∎

As in [9], [10], and [40], we will rely on the fact that, with very high probability, the norm of the noise $\boldsymbol{z}$ is not much larger than $\sqrt{n\sigma_z^2}$. The probability that the effective noise $\boldsymbol{z}$ is 'large' is exponentially small in n. The proof of this result is given below.

Lemma 4: Let $\boldsymbol{x}$ be uniformly distributed in $\mathcal{V}(\Lambda_c)$ and $\delta > 0$ be any positive number. Then

$$\mathrm{P}\left(\|\boldsymbol{z}\|^2 > n\sigma_z^2(1+\delta)\right) \leq e^{-\frac{n(\delta - \ln(1+\delta))}{2}} + e^{-\frac{n\sigma^2\delta^2}{4}}. \tag{32}$$

Proof: We will prove (32) for every fixed realization of $\boldsymbol{x}$ in $\mathcal{V}(\Lambda_c)$, which shows that the statement of the lemma is true for any distribution of $\boldsymbol{x}$ on $\mathcal{V}(\Lambda_c)$. In the rest of the proof we will assume that $\boldsymbol{x} \in \mathcal{V}(\Lambda_c)$ is an arbitrary fixed vector and $\boldsymbol{n}$ is Gaussian distributed. Using $\|\boldsymbol{x}\|^2 \leq r_{\mathrm{cov}}^2(\Lambda_c) = n$, we have

$$\|\boldsymbol{z}\|^2 = \|\alpha\boldsymbol{n} - (1-\alpha)\boldsymbol{x}\|^2$$
$$= \alpha^2\|\boldsymbol{n}\|^2 + (1-\alpha)^2\|\boldsymbol{x}\|^2 - 2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n}$$
$$\leq \alpha^2\|\boldsymbol{n}\|^2 + (1-\alpha)^2 \, n - 2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n}.$$

Hence, $\mathrm{P}\left(\|\boldsymbol{z}\|^2 > n\sigma_z^2(1+\delta)\right)$ is upper bounded by

$$\mathrm{P}\left(\alpha^2\|\boldsymbol{n}\|^2 + (1-\alpha)^2 \, n - 2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n} > n\sigma_z^2(1+\delta)\right).$$

From the definition (27) of σ_z^2, we have $n\sigma_z^2(1+\delta) = n\alpha^2\sigma^2(1+\delta) + n(1-\alpha)^2(1+\delta)$. Hence, the above upper bound corresponds to the event

$$\alpha^2\|\boldsymbol{n}\|^2 + (1-\alpha)^2 n - 2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n}$$
$$> n\alpha^2\sigma^2(1+\delta) + n(1-\alpha)^2(1+\delta). \tag{33}$$

The event (33) occurs *only if* at least one of the following two events occurs

$$\mathcal{E}_A: \alpha^2\|\boldsymbol{n}\|^2 > n\alpha^2\sigma^2(1+\delta), \quad \text{or}$$
$$\mathcal{E}_B: (1-\alpha)^2n - 2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n} > n(1-\alpha)^2(1+\delta). \tag{34}$$

Therefore,

$$\mathrm{P}(\|\boldsymbol{z}\|^2 > n\sigma_z^2(1+\delta)) \leq \mathrm{P}(\mathcal{E}_A \cup \mathcal{E}_B) \leq \mathrm{P}(\mathcal{E}_A) + \mathrm{P}(\mathcal{E}_B).$$

We will now individually upper bound $\mathrm{P}(\mathcal{E}_A)$ and $\mathrm{P}(\mathcal{E}_B)$, and thereby complete the proof.

A rearrangement of terms in (34) yields $\mathrm{P}(\mathcal{E}_A) = \mathrm{P}\left(\|\frac{1}{\sigma}\boldsymbol{n}\|^2 > n(1+\delta)\right)$. This is the probability that a Gaussian vector with unit variance per dimension lies outside the sphere of squared radius $n(1+\delta)$. The following is a well known upper bound on this probability (see [47])

$$\mathrm{P}(\mathcal{E}_A) \leq e^{-\frac{n(\delta - \ln(1+\delta))}{2}}.$$

The event $\mathcal{E}_B$ is equivalent to $-2\alpha(1-\alpha)\boldsymbol{x}^\mathsf{T}\boldsymbol{n} > n(1-\alpha)^2\delta$. Using $\alpha = 1/(1+\sigma^2)$, we can show that this is same as $\boldsymbol{x}^\mathsf{T}\boldsymbol{n} < -n\delta\sigma^2/2$. Since $\boldsymbol{x}^\mathsf{T}\boldsymbol{n}$ is a zero mean Gaussian random variable with variance $\sigma^2\|\boldsymbol{x}\|^2$, we have

$$\mathrm{P}(\mathcal{E}_B) = \mathrm{P}\left(\boldsymbol{x}^\mathsf{T}\boldsymbol{n} < -\frac{n\delta\sigma^2}{2}\right) = \mathcal{Q}\left(\frac{n\sigma\delta}{2\|\boldsymbol{x}\|}\right),$$

where $\mathcal{Q}(\cdot)$ is the Gaussian tail function. Using $\|\boldsymbol{x}\| \leq r_{\mathrm{cov}}(\Lambda_{\mathrm{c}}) = \sqrt{n}$ and the Chernoff bound $\mathcal{Q}(y) \leq \exp(-y^2/2)$, we arrive at

$$\mathrm{P}(\mathcal{E}_B) \leq e^{-\frac{n\sigma^2\delta^2}{4}}.$$

This completes the proof. ∎

B. Error Probability at a Single Receiver

In this subsection we derive an upper bound on the decoding error probability $\mathrm{P}_{\boldsymbol{S}}$ at a receiver $(\boldsymbol{S}, \sigma^2)$ when averaged over the ensemble of lattice codes generated by choosing $\boldsymbol{G}$ uniformly over $\mathbb{F}_p^{n \times K\ell}$ and $\boldsymbol{d}$ uniformly over $\mathcal{V}(\Lambda_{\mathrm{c}})$.

The following result from [8], known as the *Crypto lemma*, captures an important characteristic of random dithering.

Lemma 5 [8]: Let $\boldsymbol{t} \in \mathcal{V}(\Lambda_{\mathrm{c}})$ be any random vector. If $\boldsymbol{d}$ is independent of $\boldsymbol{t}$ and is uniformly distributed over $\mathcal{V}(\Lambda_{\mathrm{c}})$, then $\boldsymbol{x} = [\boldsymbol{t} - \boldsymbol{d}] \bmod \Lambda_{\mathrm{c}}$ is independent of $\boldsymbol{t}$ and uniformly distributed over $\mathcal{V}(\Lambda_{\mathrm{c}})$.

The property that the transmit vector $\boldsymbol{x}$ is statistically independent of $\boldsymbol{t}$ implies that the effective noise $\boldsymbol{z} = \alpha\boldsymbol{n} - (1-\alpha)\boldsymbol{x}$ is independent of the transmit message. This facilitates the error probability analysis through the observation that the error event (30) is statistically independent of $\tilde{\boldsymbol{t}}$.

For distinct messages $\tilde{\boldsymbol{w}}$ to be mapped to distinct points $\tilde{\boldsymbol{t}}$, we require that $\boldsymbol{G}\boldsymbol{A}_{\boldsymbol{S}}$ be full rank. Since $\boldsymbol{A}_{\boldsymbol{S}}$ is full rank, this is same as requiring that $\boldsymbol{G}$ be full rank. Apart from the event $\mathcal{E}: Q_{\Lambda_{\boldsymbol{S}}}(\boldsymbol{z}) \in \Lambda_{\boldsymbol{S}} \backslash \Lambda_{\mathrm{c}}$, we assume that the decoder declares an error whenever the event

$$\mathcal{G}: \mathrm{rank}(\boldsymbol{G} < K\ell)$$

occurs. Hence, the error probability $\mathrm{P}_{\boldsymbol{S}}$ at the receiver $(\boldsymbol{S}, \sigma^2)$ satisfies

$$\mathrm{P}_{\boldsymbol{S}} = \mathrm{P}(\mathcal{G} \cup \mathcal{E}) \leq \mathrm{P}(\mathcal{G}) + \mathrm{P}(\mathcal{E}). \tag{35}$$

From (18), we already know that $\mathrm{P}(\mathcal{G})$ is exponentially small in n.

Using the given design tolerance ϵ, we set $\delta = 2^{\frac{\epsilon}{2}} - 1$, which is positive if $\epsilon > 0$. Let $r_z = \sqrt{n(1+\delta)\sigma_z^2}$ be the radius of the typical noise vector and $\mathcal{B}_{r_z} = \mathcal{B}(\boldsymbol{0}, r_z)$. Then,

$$\mathrm{P}(\mathcal{E}) = \mathrm{P}(\boldsymbol{z} \in \mathcal{B}_{r_z}) \mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) + \mathrm{P}(\boldsymbol{z} \notin \mathcal{B}_{r_z}) \mathrm{P}(\mathcal{E}|\boldsymbol{z} \notin \mathcal{B}_{r_z})$$
$$\leq \mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) + \mathrm{P}(\boldsymbol{z} \notin \mathcal{B}_{r_z}). \tag{36}$$

Lemma 4 provides an exponential upper bound on $\mathrm{P}(\boldsymbol{z} \notin \mathcal{B}_{r_z})$. In the following theorem we show that $\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z})$ is also exponentially small in n. The proof of this result uses the technique of [9] and [10] to bound the number of lattice points lying in an n-dimensional ball.

Theorem 2: For any receiver $(\boldsymbol{S}, \sigma^2)$ with $\frac{1}{2} \log\left(1 + \frac{1}{\sigma^2}\right) > (R + \epsilon)(K - \mathrm{rank}(\boldsymbol{S}))$, and for all large enough n,

$$\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) \leq 2^{-\frac{n\epsilon}{4}}$$

when averaged over the ensemble of random lattice codes.

Proof: From (30), we note that the decoder is in error when $\boldsymbol{z}$ is closer to some coset $\boldsymbol{t}' + \Lambda_{\mathrm{c}}$, with $\boldsymbol{t}' \in \Lambda_{\boldsymbol{S}}/\Lambda_{\mathrm{c}}$ and $\boldsymbol{t}' \neq \boldsymbol{0}$, than any point in Λ_{c}. The number of competing cosets

is $|\Lambda_{\boldsymbol{S}}/\Lambda_{\mathrm{c}} \backslash \{\boldsymbol{0}\}| = p^{(K-M)\ell} - 1$, and we index them using the non-zero vectors $\boldsymbol{w}' \in \mathbb{F}_p^{(K-M)\ell} \backslash \{\boldsymbol{0}\}$. To each $\boldsymbol{w}'$, we associate the coset corresponding to the coset leader

$$\boldsymbol{t}' = \left[\boldsymbol{B}_{\mathrm{c}} p^{-1} g(\boldsymbol{G}\boldsymbol{A}_{\boldsymbol{S}}\boldsymbol{w}')\right] \bmod \Lambda_{\mathrm{c}}. \tag{37}$$

Since $\boldsymbol{G}$ is random, the coset leader $\boldsymbol{t}'$ associated with a given $\boldsymbol{w}'$ is a random vector. Given that $\boldsymbol{z} \in \mathcal{B}_{r_z}$ and $\boldsymbol{0} \in \Lambda_{\mathrm{c}}$, the Euclidean distance between $\boldsymbol{z}$ and Λ_{c} is at the most $\|\boldsymbol{z} - \boldsymbol{0}\| \leq r_z$. Hence, for an error event to occur, there must exist a coset $\boldsymbol{t}' + \Lambda_{\mathrm{c}}$ at a distance less than r_z from $\boldsymbol{z}$, i.e., $|(\boldsymbol{t}' + \Lambda_{\mathrm{c}}) \cap \mathcal{B}(\boldsymbol{z}, r_z)| \neq 0$. Indexing the cosets by $\boldsymbol{w}'$, we upper bound $\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z})$ using (38) given in the top of the next page.

The last inequality in (38) follows from the observation

$$\mathbf{1}\left\{\left|(\boldsymbol{t}' + \Lambda_{\mathrm{c}}) \cap \mathcal{B}(\boldsymbol{z}, r_z)\right| \neq 0\right\} \leq \left|(\boldsymbol{t}' + \Lambda_{\mathrm{c}}) \cap \mathcal{B}(\boldsymbol{z}, r_z)\right|,$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Note that the expectation operation in (38) is with respect to the random vector $\boldsymbol{t}'$ as well as the effective noise $\boldsymbol{z}$.

The matrix $\boldsymbol{A}_{\boldsymbol{S}}$ has full column rank, and hence, $\boldsymbol{A}_{\boldsymbol{S}}\boldsymbol{w}' \neq \boldsymbol{0}$ for every $\boldsymbol{w}' \neq \boldsymbol{0}$. Using (37) and applying Lemma 2, we see that $\boldsymbol{t}'$ is uniformly distributed in $p^{-1}\Lambda_{\mathrm{c}}/\Lambda_{\mathrm{c}} = p^{-1}\Lambda_{\mathrm{c}} \cap \mathcal{V}(\Lambda_{\mathrm{c}})$. Further, from Lemma 5 the distribution of $\boldsymbol{t}'$ is independent of $\boldsymbol{z}$. Hence, the probability mass function of $\boldsymbol{t}'$ equals $|\left(p^{-1}\Lambda_{\mathrm{c}}\right)/\Lambda_{\mathrm{c}}|^{-1} = p^{-n}$ over every element of the set $\left(p^{-1}\Lambda_{\mathrm{c}}\right)/\Lambda_{\mathrm{c}}$. Using this result, we further upper bound $\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z})$ as in (39) in the top of next page, where the last equality follows from the fact that the set of cosets $\left\{\boldsymbol{a} + \Lambda_{\mathrm{c}}|\boldsymbol{a} \in p^{-1}\Lambda_{\mathrm{c}}/\Lambda_{\mathrm{c}}\right\}$ form a partition of $p^{-1}\Lambda_{\mathrm{c}}$. Since the number of competing $\boldsymbol{w}'$ in (39) is less than $p^{(K-M)\ell}$, and $\left|p^{-1}\Lambda_{\mathrm{c}} \cap \mathcal{B}(\boldsymbol{z}, r_z)\right| = |\Lambda_{\mathrm{c}} \cap \mathcal{B}(p\boldsymbol{z}, pr_z)|$, we obtain

$$\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) \leq p^{-n} p^{(K-M)\ell} \mathbb{E}\left(|\Lambda_{\mathrm{c}} \cap \mathcal{B}(p\boldsymbol{z}, pr_z)| \,\big|\, \boldsymbol{z} \in \mathcal{B}_{r_z}\right).$$

Using Lemma 3, we bound the number of lattice points inside the ball $\mathcal{B}(p\boldsymbol{z}, pr_z)$, and obtain

$$\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) \leq p^{-n} p^{(K-M)\ell} \frac{V_n}{\mathrm{Vol}(\Lambda_{\mathrm{c}})} \left(r_{\mathrm{cov}}(\Lambda_{\mathrm{c}}) + pr_z\right)^n.$$

Using the bounds $p^{(K-M)\ell} \leq 2^{nR(K-M)}$, from (13); $\mathrm{Vol}(\Lambda_{\mathrm{c}}) \geq V_n n^{\frac{n}{2}} 2^{-\frac{n}{4}}$, from (14); $\sigma_z \leq 2^{-(R+\epsilon)(K-M)}$, from (28); and the relations $r_{\mathrm{cov}}(\Lambda_{\mathrm{c}}) = \sqrt{n}$, $r_z = \sqrt{n(1+\delta)\sigma_z^2}$, and $1 + \delta = 2^{\frac{\epsilon}{2}}$, we obtain the sequence of equalities and upper bounds leading to (40) shown in the next page.

Since $K - M \geq 1$, we have $2^{n\epsilon(K-M)} \geq 2^{n\epsilon}$, and hence, the upper bound (40) can be further relaxed as

$$\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) \leq 2^{-\frac{n\epsilon}{2}} \left(\frac{1}{p\, 2^{(R+\epsilon)(K-M)}} + 1\right)^n.$$

Using the inequality (12), which immediately follows from the choice of the prime integer p, we have

$$\mathrm{P}(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_z}) \leq 2^{-\frac{n\epsilon}{2}} 2^{\frac{n\epsilon}{4}} = 2^{-\frac{n\epsilon}{4}}.$$

Note that this upper bound holds for all sufficiently large values of n. ∎

$$P(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}}) \leq P\left(\bigcup_{\boldsymbol{w}' \in \mathbb{F}_p^{(K-M)\ell} \setminus \{\boldsymbol{0}\}} \left\{ \left| (\boldsymbol{t}' + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z}) \right| \neq 0 \right\} \Big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right)$$

$$\leq \sum_{\boldsymbol{w}' \in \mathbb{F}_p^{(K-M)\ell} \setminus \{\boldsymbol{0}\}} P\left(\left| (\boldsymbol{t}' + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z}) \right| \neq 0 \Big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right)$$

$$\leq \sum_{\boldsymbol{w}' \in \mathbb{F}_p^{(K-M)\ell} \setminus \{\boldsymbol{0}\}} \mathbb{E}\left(\left| (\boldsymbol{t}' + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z}) \right| \Big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right). \tag{38}$$

$$P(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}}) \leq \sum_{\boldsymbol{w}'} \sum_{\boldsymbol{a} \in p^{-1}\Lambda_c / \Lambda_c} P(\boldsymbol{t}' = \boldsymbol{a}) \, \mathbb{E}\left(|(\boldsymbol{a} + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z})| \big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right)$$

$$= \sum_{\boldsymbol{w}'} \sum_{\boldsymbol{a} \in p^{-1}\Lambda_c / \Lambda_c} p^{-n} \mathbb{E}\left(|(\boldsymbol{a} + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z})| \big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right)$$

$$= p^{-n} \sum_{\boldsymbol{w}'} \sum_{\boldsymbol{a} \in p^{-1}\Lambda_c / \Lambda_c} \mathbb{E}\left(|(\boldsymbol{a} + \Lambda_c) \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z})| \big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right) = p^{-n} \sum_{\boldsymbol{w}'} \mathbb{E}\left(\left| p^{-1}\Lambda_c \cap \mathcal{B}(\boldsymbol{z}, r_\boldsymbol{z}) \right| \big| \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}} \right). \tag{39}$$

$$P(\mathcal{E}|\boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}}) \leq p^{-n} \, 2^{nR(K-M)} \frac{V_n}{V_n n^{\frac{n}{2}} 2^{-\frac{n\epsilon}{4}}} \left(\sqrt{n} + p\sqrt{n(1+\delta)\sigma_\boldsymbol{z}^2} \right)^n$$

$$= \frac{2^{nR(K-M)} 2^{\frac{n\epsilon}{4}}}{n^{\frac{n}{2}}} \left(\frac{\sqrt{n}}{p} + \sqrt{n(1+\delta)\sigma_\boldsymbol{z}^2} \right)^n = \frac{2^{nR(K-M)} 2^{\frac{n\epsilon}{4}}}{n^{\frac{n}{2}}} \left(\frac{\sqrt{n}}{p} + \sqrt{n(1+\delta)} 2^{-(R+\epsilon)(K-M)} \right)^n$$

$$= \frac{2^{nR(K-M)} 2^{\frac{n\epsilon}{4}}}{n^{\frac{n}{2}}} n^{\frac{n}{2}} (1+\delta)^{\frac{n}{2}} \left(\frac{1}{p\sqrt{(1+\delta)}} + 2^{-(R+\epsilon)(K-M)} \right)^n$$

$$\leq 2^{nR(K-M)} 2^{\frac{n\epsilon}{4}} 2^{\frac{n\epsilon}{4}} \left(\frac{1}{p} + 2^{-(R+\epsilon)(K-M)} \right)^n \leq \frac{2^{nR(K-M)} 2^{\frac{n\epsilon}{2}}}{2^{n(R+\epsilon)(K-M)}} \left(\frac{1}{p \, 2^{(R+\epsilon)(K-M)}} + 1 \right)^n$$

$$= \frac{2^{nR(K-M)} 2^{\frac{n\epsilon}{2}}}{2^{nR(K-M)} \, 2^{n\epsilon(K-M)}} \left(\frac{1}{p \, 2^{(R+\epsilon)(K-M)}} + 1 \right)^n. \tag{40}$$

We will now combine the result of Theorem 2 with (35) and (36), and upper bound the error probability $P_\boldsymbol{S}$ at the receiver $(\boldsymbol{S}, \sigma^2)$ as

$$P_\boldsymbol{S} \leq P(\mathcal{G}) + P(\boldsymbol{z} \notin \mathcal{B}_{r_\boldsymbol{z}}) + P(\mathcal{E} \mid \boldsymbol{z} \in \mathcal{B}_{r_\boldsymbol{z}}).$$

Using Theorem 2, Lemma 4 and (18), we obtain

$$P_\boldsymbol{S} \leq 2^{-\frac{n}{2}} + e^{-\frac{n(\delta - \ln(1+\delta))}{2}} + e^{-\frac{n\sigma^2\delta^2}{4}} + 2^{-\frac{n\epsilon}{4}}, \tag{41}$$

for sufficiently large n. Let $\sigma_{\min} > 0$ be the least noise standard deviation σ among the finitely many receivers in the multicast channel. Then we have $\sigma^2\delta^2/4 \geq \sigma_{\min}^2\delta^2/4$. Also, $\delta - \ln(1 + \delta) > 0$ as long as $\delta = 2^{\frac{\epsilon}{2}} - 1$ is positive. Consequently, the parameter

$$\varepsilon = \min\left\{ \frac{1}{2}, \log e \left(\frac{\delta - \ln(1+\delta)}{2} \right), \log e \left(\frac{\sigma_{\min}^2 \delta^2}{4} \right), \frac{\epsilon}{4} \right\}$$

is positive, and the value of each of the terms on the right-hand side of (41) is at most $2^{-n\varepsilon}$. Hence the error probability at the receiver $(\boldsymbol{S}, \sigma^2)$ can be upper bounded as

$$P_\boldsymbol{S} \leq 4 \cdot 2^{-n\varepsilon}, \tag{42}$$

for all sufficiently large n. We remark that the minimum required value of n for this upper bound to hold

depends only ϵ, and is independent of the side information matrix $\boldsymbol{S}$.

C. Completing the Proof of the Main Theorem

The bound (42) shows that the error probability for a fixed side information matrix $\boldsymbol{S}$, averaged over the random code ensemble, tends to 0 as the code dimension increases. Hence, there exists a choice of lattice code (which is chosen for the given side information matrix $\boldsymbol{S}$) with a small error probability at this receiver. We want to prove a slightly stronger result, viz., there exists a lattice code such that the decoding error probability for every possible side information matrix $\boldsymbol{S}$ is small as long as the receiver SNR is large enough. In order to prove this result, we consider a hypothetical multicast network that consists of one receiver for each possible choice of the matrix $\boldsymbol{S}$. Note that two distinct values of the matrix $\boldsymbol{S}$ that have identical row space constitute equivalent receiver side information configurations. Hence, it is enough to consider a multicast channel that consists of one receiver corresponding to each possible subspace of $\mathbb{F}_p^K$, where the dimension of the subspace can be between 0 and $K - 1$. A subspace of dimension M, $0 \leq M \leq K - 1$, can be mapped to an $M \times K$ matrix whose rows form a basis of the subspace. This map embeds the set $\mathcal{S}$ of all non-equivalent choices of side

information matrix S into $\cup_{M=0}^{K-1} \mathbb{F}_p^{M \times K}$, which is the set of all matrices over $\mathbb{F}_p$ with K columns and at the most $K-1$ rows. Hence, the number of receivers $|\mathcal{S}|$ can be upper bounded as

$$|\mathcal{S}| \leq \sum_{M=0}^{K-1} \left| \mathbb{F}_p^{M \times K} \right| = \sum_{M=0}^{K-1} p^{MK} \leq K p^{K^2}. \quad (43)$$

We assume that each receiver (S, σ^2), $S \in \mathcal{S}$, satisfies the lower bound (10) on SNR and outputs an estimated message vector $\hat{w}(S)$ using its own channel observation. We say that the multicast network is in error if any of the receivers commits a decoding error. Using a union bound argument and the upper bounds (42) and (43), we see that the network error probability P_{net} averaged over the random ensemble of lattice codes satisfies

$$\begin{aligned} P_{\text{net}} &= P(\text{network error}) = P\left(\bigcup_{S \in \mathcal{S}} \{\hat{w}(S) \neq w\} \right) \\ &\leq \sum_{S \in \mathcal{S}} P\left(\{\hat{w}(S) \neq w\}\right) = \sum_{S \in \mathcal{S}} P_S \\ &\leq 4K \, p^{K^2} 2^{-n\varepsilon}, \end{aligned} \quad (44)$$

which tends to 0 as n becomes arbitrarily large. Hence, for every sufficiently large n, there exists a lattice code such that the network error probability is as small as desired. In particular, this implies that there exists a choice of lattice code such that the decoding error probability at every receiver (S, σ^2), $S \in \mathcal{S}$, is simultaneously small. This completes the proof of the main theorem.

D. Corollaries

1) Almost All Lattice Codes Are Good: Using standard arguments based on Markov inequality [8], [9], [46], we show that almost all codes from the random lattice code ensemble yield a small error probability. In order to prove this, it is sufficient to show that for almost all lattice codes the network error probability is small over the hypothetical multicast channel that consists of one receiver for each possible side information matrix.

For a given dimension n, all the lattice codes in the random code ensemble use the same coarse lattice Λ_c, but differ in the choice of the fine lattice Λ and/or the dither vector d. Let

$$X(\Lambda, d) = P(\text{network error} | \Lambda, d)$$

denote the network error probability for a given choice of Λ, d in the hypothetical multicast channel. If Λ and d are chosen randomly, then X is a random variable. From (44), we know that the expected value of X, which is equal to the average network error rate P_{net}, is small. Suppose we want a lower bound on the fraction of random codes with error probability at the most $2^{-\frac{n\varepsilon}{2}}$. Using Markov inequality, we have

$$P\left(X > 2^{-\frac{n\varepsilon}{2}}\right) \leq \frac{\mathbb{E}(X)}{2^{-\frac{n\varepsilon}{2}}} \leq \frac{4K \, p^{K^2} 2^{-n\varepsilon}}{2^{-\frac{n\varepsilon}{2}}} = 4K \, p^{K^2} 2^{-\frac{n\varepsilon}{2}}.$$

It follows that, asymptotically in n, for almost all choices of the fine lattice Λ and dither vector d, the resulting lattice code

$(\Lambda - d)/\Lambda_c$ provides an exponentially small error probability in the multicast channel, i.e.,

$$\lim_{n \to \infty} P\left(X \leq 2^{-\frac{n\varepsilon}{2}}\right) = 1.$$

2) Goodness in Single-User AWGN Channel: Our model of multicast channel includes as a special case the single-transmitter single-receiver AWGN channel with no side information at the receiver, i.e., number of messages $K = 1$, side information matrix S is the empty matrix and rank$(S) = M = 0$. The decoder for this receiver uses the $\ell \times \ell$ identity matrix for A_S and the all zero vector for v, see (21). Specializing the main theorem for a single receiver with $M = 0$, we immediately deduce that the ensemble of random lattice codes achieves the capacity of the single-user AWGN channel and hence arrive at Corollary 1.

It is well known that (nested) lattice codes, and lattice constellations in general, can achieve the capacity of the point-to-point AWGN channel [7]–[11]. Our corollary to the main theorem provides an alternate proof of this result which is based only on simple counting arguments.

The proof technique presented in this paper relies on lattices obtained by applying Construction A to random linear codes over a large enough prime field $\mathbb{F}_p$. This technique was introduced by Loeliger [40] and used in [8]–[10] to prove the goodness of lattice codes in AWGN channel. Each of these results requires a different choice of the prime p and places different requirements on the characteristics of the coarse lattice Λ_c. The following are some of the properties that have been used in the literature:

- *Rogers-good:* the ratio of covering radius $r_{\text{cov}}(\Lambda_c)$ to the effective radius $r_{\text{eff}}(\Lambda_c)$ of the lattice must be close to 1, see (8). Such a lattice is also said to be *good for covering.*
- *MSE-good:* the value of the lattice parameter

$$\frac{1}{n \operatorname{Vol}(\Lambda_c)^{1+\frac{2}{n}}} \int_{\mathcal{V}(\Lambda_c)} \|x\|^2 dx,$$

known as the *normalized second moment*, is close to $1/2\pi e$, see [45]. Every Rogers-good lattice is also MSE-good, and hence, this is a weaker requirement.
- *Poltyrev-good:* such a lattice, when used as an infinite constellation, achieves the capacity of an AWGN channel in which the transmitter has no power constraints [45], [47]. These lattices are resilient against additive white Gaussian noise.

The achievability result of [8] requires Λ_c to be simultaneously Rogers-good and Poltyrev-good, and uses $p = 2^{nR}$, i.e., the prime field $\mathbb{F}_p$ used for Construction A varies with the dimension of the lattice code and the size of the field increases exponentially in n. The random code ensemble of [9] uses an MMSE-good lattice for Λ_c, lets p grow as $n^{1.5}$, and can accommodate a wide class of channel noise statistics, including white Gaussian noise. The code construction of [10] requires p to be at least $n^{0.5}$, needs no dithering operation, i.e., uses $d = 0$, but is known to achieve capacity only if SNR > 1. In comparison, our proof method uses a fixed (albeit large) value of p and holds for any SNR > 0, while requiring that Λ_c be Rogers-good.

V. CONCLUSION

We have showed that lattice codes are optimal for common message broadcast in Gaussian channels where receivers have side information in the form of linear combinations of source messages. We used random lattice ensembles obtained by applying Construction A to linear codes over appropriately large prime fields $\mathbb{F}_p$. The lower bound $p \geq \max\left\{2^{2KR}, (2^{\frac{\epsilon}{4}}-1)^{-1}2^{-R}\right\}$ on the value of p does not necessarily pose a limitation in communication applications. For instance, in the relay network of Example 1, the first phase of the protocol, namely compute-and-forward [15], only requires that $\frac{n}{p} \to 0$ as $n \to \infty$, which can be met by our scheme by varying p with the dimension n: for instance, by choosing p to be the smallest prime greater than or equal to n^β for a fixed $\beta > 1$. This will also ensure that the inequality $p \geq \max\left\{2^{2KR}, (2^{\frac{\epsilon}{4}}-1)^{-1}2^{-R}\right\}$ holds for all sufficiently large values of n. Similarly, with Example 2, where the broadcast signal supplements a wired multicast network, it is known that wireline network codes meeting the max-flow bound exist over every large enough finite field [26]. Hence, we can choose p to be sufficiently large to simultaneously optimize both the wired and wireless parts of the hybrid network. On the other hand, designing lattice strategies for a fixed small size of the finite field, especially sizes that are powers of two, may have greater practical significance.

The capacity of the Gaussian broadcast channel with receiver side information under general message demands, such as with private message requests, is known only for some special cases [29], [48], [49]. The proofs for achievability in these cases utilize ensembles of codebooks generated using the Gaussian distribution together with dirty-paper and superposition coding. It will be interesting to examine if the lattice structure of the codes proposed in this paper can be exploited to derive new capacity results beyond the known cases.

APPENDIX
CAPACITY OF THE COMMON MESSAGE GAUSSIAN BROADCAST CHANNEL WITH CODED SIDE INFORMATION

Consider the problem setup with a single transmitter and N receivers $(\boldsymbol{S}_1, \sigma_1^2), \ldots, (\boldsymbol{S}_N, \sigma_N^2)$ as described in Section II-A. We now provide a sketch of the proof that C, defined in (4), is the capacity of this channel.

A. Converse

Suppose there exists a coding scheme that achieves rate R in the multicast channel with vanishing decoding error probability at all the receivers. Let the scheme transmit one realization of $(w_1, \ldots, w_K)$ for every κ channel uses, i.e., $R = \frac{1}{\kappa} \log_2 p$. From (1) the conditional entropy of each realization of $(w_1, \ldots, w_K)$ at the ith receiver $(\boldsymbol{S}_i, \sigma_i^2)$, given the corresponding coded side information, is $(K - M_i) \log_2 p$. The per-channel use conditional entropy of the message is thus

$$\frac{(K - M_i)\log_2 p}{\kappa} = (K - M_i)R.$$

In order to guarantee reliable communication it is necessary that the mutual information between the channel input at the transmitter and the channel output at the ith receiver be greater than the conditional entropy $(K - M_i)R$. Since the input power is constrained to be at the most 1 and the noise variance at the ith receiver is σ_i^2, the maximum mutual information is $\frac{1}{2}\log_2\left(1 + \frac{1}{\sigma_i^2}\right)$, and hence we have

$$(K - M_i)R < \frac{1}{2}\log_2\left(1 + \frac{1}{\sigma_i^2}\right), \quad \text{or equivalently,}$$

$$R < \frac{1}{(K - M_i)} \cdot \frac{1}{2}\log_2\left(1 + \frac{1}{\sigma_i^2}\right).$$

Considering all the N receivers we immediately deduce that $R < C$.

B. Achievability

The proof of achievability closely follows the proof of [17, Th. 6] and the standard textbook argument used for the achievability of the capacity of single-user AWGN channel. Let $\epsilon > 0$ be any constant. For a given code length n choose the message length ℓ as the largest integer such that the rate $R = \frac{\ell}{n}\log_2 p$ satisfies

$$R < \min_{i \in \{1, \ldots, N\}} \frac{1}{(K - M_i)} \cdot \frac{1}{2}\log_2\left(1 + \frac{1 - \epsilon}{\sigma_i^2}\right) - 3\epsilon.$$

As $n \to \infty$, it is straightforward to show that R converges to the right-hand side of the above inequality. For each of the 2^{nKR} message vectors $\boldsymbol{w} = (\boldsymbol{w}_1^{\mathsf{T}}, \ldots, \boldsymbol{w}_K^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{F}_p^{K\ell}$, associate a codeword $\boldsymbol{x}(\boldsymbol{w}) \in \mathbb{R}^n$ each of whose components are generated independently using the Gaussian distribution with zero mean and variance $(1 - \epsilon)$. These 2^{nKR} vectors constitute the randomly-generated n-dimensional codebook $\mathcal{X}$.

Encoding: If the source message is $\boldsymbol{w}$, the transmitter broadcasts the vector $\boldsymbol{x}(\boldsymbol{w})$ over n channel uses.

Decoding: Consider the ith receiver $(\boldsymbol{S}_i, \sigma_i^2)$ that observes the channel output $\boldsymbol{y}_i$ and the side information $\sum_{k=1}^{K} s_{m,k}^{(i)} \boldsymbol{w}_k = \boldsymbol{u}_m^{(i)}$, $m = 1, \ldots, M_i$, where $\boldsymbol{S}_i = [s_{m,k}^{(i)}]$. As in (3), the receiver determines the subcode $\mathcal{X}_{\text{sub}} = \mathcal{X}(\boldsymbol{S}_i, \boldsymbol{u}_1^{(i)}, \cdots, \boldsymbol{u}_{M_i}^{(i)})$ of the codebook $\mathcal{X}$ that corresponds to the set of all message vectors $\boldsymbol{w}$ which are consistent with the observed coded side information. Among the codewords in $\mathcal{X}_{\text{sub}}$, the decoder chooses the vector that is jointly (weakly) ϵ-typical with $\boldsymbol{y}_i$. If there exists a unique such codeword $\boldsymbol{x}(\hat{\boldsymbol{w}})$ that additionally satisfies the power constraint $\|\boldsymbol{x}(\hat{\boldsymbol{w}})\|^2 \leq n$, the receiver declares $\hat{\boldsymbol{w}}$ as the decoded message. Otherwise the receiver declares a decoding error.

Given that $\mathcal{X}_{\text{sub}}$ consists of $2^{n(K-M_i)R}$ vectors generated independently using the Gaussian distribution with zero mean and variance $(1 - \epsilon)$ and

$$(K - M_i)R < \frac{1}{2}\log_2\left(1 + \frac{1 - \epsilon}{\sigma_i^2}\right) - 3\epsilon,$$

it is routine to show that the probability of decoding error at the ith receiver, averaged over the ensemble

of codebooks, decays exponentially with code length n [50, Proof of Theorem 10.1.1]. It follows that the probability that any of the N receivers commits a decoding error is also exponentially small in n. Hence, there exists at least one codebook $\mathcal{X}$ that transmits each message at rate R with the decoding error probability at all the receivers as small as desired. Letting $n \to \infty$ and $\epsilon \to 0$, we observe that any rate $R < C$ is achievable.

ACKNOWLEDGMENT

REFERENCES

[1] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.

[2] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[3] B. Nazer and M. Gastpar, "The case for structured random codes in network capacity theorems," *Trans. Emerg. Telecommun. Technol.*, vol. 19, no. 4, pp. 455–474, 2008.

[4] G. D. Forney, Jr., "Coset codes. I. Introduction and geometrical classification," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 5, pp. 1123–1151, Sep. 1988.

[5] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York, NY, USA: Springer-Verlag, 1999.

[6] R. de Buda, "The upper error bound of a new near-optimal code," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 4, pp. 441–445, Jul. 1975.

[7] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, Jan. 1998.

[8] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[9] O. Ordentlich and U. Erez, "A simple proof for the existence of 'good' pairs of nested lattices," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4439–4453, Aug. 2016.

[10] N. Di Pietro, "On infinite and finite lattice constellations for the additive white Gaussian noise channel," M.S. Theses, Dept. Comput. Sci., Univ. Bordeaux, Bordeaux, France, Jan. 2014. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01135575

[11] C. Ling and J.-C. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[12] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[13] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.

[14] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[15] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[16] C. Feng, D. Silva, and F. R. Kschischang, "An algebraic approach to physical-layer network coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7576–7596, Nov. 2013.

[17] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

[18] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.

[19] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *Proc. 49th IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2008, pp. 823–832.

[20] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, Jul. 2010.

[21] K. W. Shum, M. Dai, and C. W. Sung, "Broadcasting with coded side information," in *Proc. IEEE 23rd Int. Symp Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 89–94.

[22] N. Lee, A. G. Dimakis, and R. W. Heath, "Index coding with coded side-information," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 319–322, Mar. 2015.

[23] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *Proc. 17th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 3. Mar./Apr. 1998, pp. 1257–1264.

[24] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[25] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[26] R. W. Yeung, *Information Theory and Network Coding*. New York, NY, USA: Springer, 2008.

[27] L.-L. Xie, "Network coding and random binning for multi-user channels," in *Proc. 10th Can. Workshop Inf. Theory (CWIT)*, Jun. 2007, pp. 85–88.

[28] G. Kramer and S. Shamai (Shitz), "Capacity for classes of broadcast channels with receiver side information," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2007, pp. 313–318.

[29] Y. Wu, "Broadcasting when receivers know some messages *a priori*," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 1141–1145.

[30] T. J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 454–458, Jan. 2008.

[31] F. Xue and S. Sandhu, "PHY-layer network coding for broadcast channel with side information," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2007, pp. 108–113.

[32] T. Wang, S. C. Liew, and L. Shi. (2015). "A lattice approach for optimal rate-diverse wireless network coding." [Online]. Available: https://arxiv.org/abs/1509.07250v1

[33] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello, "Nested codes with multiple interpretations," in *Proc. 40th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2006, pp. 851–856.

[34] F. C. Barbosa and M. H. M. Costa, "A tree construction method of nested cyclic codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2011, pp. 302–305.

[35] Y. Ma, Z. Lin, H. Chen, and B. Vucetic, "Multiple interpretations for multi-source multi-destination wireless relay network coded systems," in *Proc. IEEE 23rd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 2253–2258.

[36] L. Natarajan, Y. Hong, and E. Viterbo, "Lattice index coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6505–6525, Dec. 2015.

[37] Y.-C. Huang, "Lattice index codes from algebraic number fields," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 2485–2489.

[38] L. Natarajan, Y. Hong, and E. Viterbo, "Index codes for the Gaussian broadcast channel using quadrature amplitude modulation," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1291–1294, Aug. 2015.

[39] L. Natarajan, Y. Hong, and E. Viterbo, "Capacity of coded index modulation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 596–600.

[40] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

[41] D. Krithivasan and S. S. Pradhan. (Jul. 2007). *A proof of the Existence of Good Nested Lattices*. [Online]. Available: http://www.eecs.umich.edu/techreports/systems/cspl/cspl-384.pdf

[42] G. D. Forney, "Multidimensional constellations. II. Voronoi constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.

[43] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[44] C. A. Rogers, "Lattice coverings of space," *Mathematika*, vol. 6, no. 1, pp. 33–39, 1959.

[45] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[46] G. D. Forney, Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41st Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2003, pp. 430–439.

[47] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 409–417, Mar. 1994.

[48] B. Asadi, L. Ong, and S. J. Johnson, "The capacity of three-receiver AWGN broadcast channels with receiver message side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 2899–2903.

[49] J. Sima and W. Chen, "Joint network and Gelfand–Pinsker coding for 3-receiver Gaussian broadcast channels with receiver message side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 81–85.

[50] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

Lakshmi Natarajan is an Assistant Professor in the Department of Electrical Engineering, Indian Institute of Technology Hyderabad. He received the B.E. degree from the College of Engineering, Guindy, in electronics and communication in 2008, and the Ph.D. degree from the Indian Institute of Science, Bangalore, in 2013. Between 2014 and 2016 he held a post-doctoral position at the Department of Electrical and Computer Systems Engineering, Monash University, Australia. His primary research interests are coding and information theory for communication systems.

Dr. Natarajan is an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He was the recipient of the Seshagiri-Kaikini Medal 2013-14 for best Ph.D. thesis, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. He was recognized as an Exemplary Reviewer by the editorial board of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2013, 2015 and 2016. He served as the Local Arrangements Co-Chair of the 2016 Australian Communications Theory Workshop, Melbourne and the 2016 Australian Information Theory School, Melbourne.

Yi Hong (S'00–M'05–SM'10) is currently a Senior Lecturer at the Department of Electrical and Computer Systems Eng., Monash University, Melbourne, Australia. She obtained her Ph.D. degree in Electrical Engineering and Telecommunications from the University of New South Wales (UNSW), Sydney, and received the NICTA-ACoRN Earlier Career Researcher Award at the Australian Communication Theory Workshop, Adelaide, Australia, 2007.

Dr. Hong was an Associate Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS and the *Transactions on Emerging Telecommunications Technologies*. She was the General Co-Chair of the IEEE Information Theory Workshop 2014, Hobart; the Technical Program Committee Chair of the Australian Communications Theory Workshop 2011, Melbourne; and the Publicity Chair at the IEEE Information Theory Workshop 2009, Sicily. She was a Technical Program Committee member for several leading IEEE conferences. Her research interests include communication theory, coding and information theory with applications to telecommunication engineering.

Emanuele Viterbo (M'95–SM'04–F'11) is currently a Professor in the ECSE Department and an Associate Dean in Graduate Research at Monash University, Melbourne, Australia. He received his Ph.D. in 1995 in Electrical Engineering, from the Politecnico di Torino, Torino, Italy. From 1990 to 1992 he was with the European Patent Office, The Hague, The Netherlands, as a patent examiner in the field of dynamic recording and error-control coding. Between 1995 and 1997 he held a post-doctoral position in the Dipartimento di Elettronica of the Politecnico di Torino. In 1997-98 he was a post-doctoral research fellow in the Information Sciences Research Center of AT&T Research, Florham Park, NJ, USA. From 1998-2005, he worked as Assistant Professor and then Associate Professor, in Dipartimento di Elettronica at Politecnico di Torino. From 2006-2009, he worked in DEIS at University of Calabria, Italy, as a Full Professor.

Prof. Viterbo is an ISI Highly Cited Researcher since 2009. He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, the *European Transactions on Telecommunications* and the *Journal of Communications and Networks*, and Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING: Special Issue on Managing Complexity in Multiuser MIMO Systems. Prof. Viterbo was awarded a NATO Advanced Fellowship in 1997 from the Italian National Research Council. His main research interests are in lattice codes for the Gaussian and fading channels, algebraic coding theory, algebraic space-time coding, digital terrestrial television broadcasting, digital magnetic recording, and irregular sampling.