Title
Variable-Length Limited Feedback Beamforming in Multiple-Antenna Fading Channels

Journal
IEEE Transactions on Information Theory, 60(11)

ISSN
0018-9448

Authors
Koyuncu, Erdem
Jafarkhani, Hamid

Publication Date
2014-11-01

DOI
10.1109/tit.2014.2357811

Peer reviewed

Variable-Length Limited Feedback Beamforming in Multiple-Antenna Fading Channels

Erdem Koyuncu, *Member*, IEEE, and Hamid Jafarkhani, *Fellow*, IEEE

***Abstract*—We study a multiple-input single-output fading channel where we would like to minimize the channel outage probability or the symbol error rate (SER) by employing beamforming via quantized channel state information at the transmitter (CSIT). We consider a variable-length limited feedback scheme where the quantized CSIT is acquired through feedback binary codewords of possibly different lengths. We design and analyze the performance of the associated variable-length quantizers (VLQs) and compare their performance with the previously-studied fixed-length quantizers (FLQs). For the outage probability performance measure, we construct VLQs that can achieve the full-CSIT performance with finite rate. Moreover, as the signal-to-noise ratio P tends to infinity, we show that VLQs can achieve the full-CSIT outage probability performance with asymptotically zero feedback rate. For the SER performance measure, we show that while the SER with full-CSIT is not achievable at any finite feedback rate, the diversity and array gains with full-CSIT can be achieved using VLQs with asymptotically zero feedback rate as $P \to \infty$. Our results show that VLQs can significantly improve upon the traditional FLQs that require infinite feedback rate to achieve the outage probability or the diversity and array gains with full-CSIT.**

***Index Terms*—Multiple antenna systems, limited feedback, channel quantization, variable-length quantization, beamforming, outage probability, error probability, diversity and array gains.**

I. Introduction

A. Limited Feedback in Multiple Antenna Systems

The availability of channel state information (CSI) at the transmitter and/or the receiver can greatly improve the performance of multiple antenna communication systems. Typically, in a point-to-point multiple antenna system with slow fading channels, the receiver can acquire the CSI via training sequences from the transmitter. Obtaining CSI at the transmitter (CSIT) is however more difficult and generally requires feedback from the receiver. A complication in this context is that the channel state itself can take any value in a multidimensional complex space. Therefore, its exact representation requires an "infinite number of feedback bits." In practice, the feedback link has a finite bandwidth, which means that only a finite number of feedback bits per channel state can be utilized for feedback. One way to model such a limited feedback scenario is to formulate it as a source coding problem. The core element of such a formulation is a (channel) quantizer that specifies (i) for each channel state, the finite sequence of feedback bits to be fed back by the receiver; and (ii) for each such sequence, the codeword (e.g. a beamforming vector) to be employed by the transmitter. One is then concerned with the design of an optimal quantizer with respect to a specific performance measure such as outage or error probability, subject to the rate constraint of the feedback link.

This work was supported in part by the NSF Award CCF-1218771.

The material in this paper was presented in part at the IEEE Data Compression Conference (DCC), Snowbird, Utah, March 20–23, 2013, and at the IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, July 7–12, 2013.

The authors are with the Center for Pervasive Communications and Computing, University of California, Irvine, Irvine CA 92697-2625 USA. Email: {ekoyuncu, hamidj}@uci.edu.

B. Related Work

A comprehensive overview on the design and performance analysis of channel quantizers for multiple antenna systems can be found in [1]. In particular, beamforming with limited feedback has been extensively studied with several different approaches some of which we summarize in the following.

Applications of Grassmannian line packings to the quantized beamforming problem have been studied in [2]–[4]. The design and analysis of limited feedback beamforming systems using vector quantizer design algorithms such as the Generalized Lloyd Algorithm can be found in [5]–[8]. In [9], the authors employ high resolution approximations of the source coding theory to analyze quantized beamforming systems. The combination of quantized beamforming with space-time coding has been studied in [10]–[12]. Multicarrier limited feedback beamforming schemes have been introduced and analyzed in [13], [14]. Random vector quantizers for multiple antenna systems have been studied in [15]–[17]. In [18], the authors design trellis-coded quantizers for cooperative beamforming systems. Capacity-optimality of quantized beamforming have been studied in [19]. The performance of quantized beamforming for distributed MISO systems has been analyzed in [20]–[24]. Limited feedback beamforming schemes for temporally/spatially correlated channels have been considered in [25]–[27]. Other approaches to quantized beamforming include [28]–[33].

The research on limited feedback is not limited to the particular quantized CSI scenarios that we have mentioned here, numerous other schemes have been devised for different transmission strategies, performance measures, and system models; we refer the interested readers to the aforementioned survey article [1].

C. Scope of the Paper

Most of the previous work on finite-rate CSI feedback has been based on fixed-length quantizers (FLQs), in which the number of feedback bits per channel state is a fixed nonnegative integer. In general, different binary codewords of different lengths can be fed back for different channel states, resulting in what is called a variable-length quantizer (VLQ). An FLQ is a special case of a VLQ, and therefore, with the same constraint on the feedback rate, we expect the achievable performance with VLQs to be better (at least not worse) than the one with FLQs.

In this work, we consider the VLQ design problem for a multiple-input single-output (MISO) system with t transmitter antennas and a short term power constraint at the transmitter. We assume a quasi-static block fading channel model in which the channel realizations vary independently from one fading block to another while within each block they remain constant. We also assume that the receiver has full CSI, while the transmitter has only partial CSI provided by the receiver via error-free and delay-free feedback channels. The partial CSI is in the form of quantized instantaneous CSI provided by a VLQ, whose "informal structure" is as described in Section I-A. We design outage-minimizing and error-minimizing VLQs for both beamforming and precoding strategies. For the error probability performance measure, we investigate the achievable diversity and array gains of the system.

The focus of this paper is specifically on MISO systems with beamforming, and we will not consider the multiple-input multiple-output (MIMO) case where the receiver also has multiple antennas. Note that if both the transmitter and the receiver has multiple antennas, beamforming in general becomes a suboptimal strategy as it can only provide a multiplexing gain of 1. Hence, in MIMO systems, one should utilize the more general precoding strategies where several independent data streams are simultaneously transmitted over the multiple transmitter antennas. Despite extensive research efforts in the last two decades, the understanding on MIMO limited feedback systems with precoding is quite limited. For example, in any quantization process, one at least hopes the quantized performance to approach the unquantized performance as the quantization resolution (the number of codewords in the quantizer codebook) increases. It is not even known whether this fundamental property holds in the case of MIMO systems with the outage probability performance measure. In other words, it is not known whether or not one can get arbitrarily close to the full-CSIT outage probability performance by using FLQs with high-enough resolution. The lack of such basic fundamental results makes the analysis of MIMO VLQs very challenging.

D. Organization of the Paper

The rest of this paper is organized as follows: We first focus on the design of outage-optimal VLQs. In Section II, we provide a formal description of the system model, the outage probability performance measure, and the variable-length channel quantizers. In Section III, we discuss how to design an optimal VLQ for a given codebook. In Section IV, we state our main results. In Section V, we discuss how to design VLQs with fast encoders/decoders and discuss some other practical issues regarding variable-length quantization. In Section VI, we provide extensions of our results to the case where non-Gaussian input distributions with finite support are used for data transmission. In Section VII, we present numerical evidence that confirms our analytical results. In Section VIII, we design and analyze VLQs that minimize the symbol error rate. Finally, in Section IX, we draw our main conclusions. Some technical proofs are provided in the appendices.

E. Notation

The symbols $o(\cdot), \omega(\cdot), O(\cdot), \Omega(\cdot)$, $\Theta(\cdot)$, $\sim$ are the standard symbols that describe the asymptotic growth of functions [34]. Also, $\|\cdot\|$ is the 2-norm, $\langle\cdot|\cdot\rangle$ is the complex inner product. $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{N}$ represent the sets of complex numbers, real numbers, and natural numbers, respectively. The set $\mathbb{C}^{t\times r}$ represents the collection of all matrices with t rows, r columns, and entries that belong to $\mathbb{C}$. $\mathrm{tr}(\mathbf{A})$ is the trace of a square matrix $\mathbf{A}$. $\mathbf{A}^T$, $\mathbf{A}^\dagger$ denote the transpose and the Hermitian transpose of $\mathbf{A}$, respectively. Let $\mathbf{A}^* = (\mathbf{A}^T)^\dagger$. $\Re(\mathbf{A}) \triangleq \frac{1}{2}(\mathbf{A}+\mathbf{A}^H)$ and $\Im(\mathbf{A}) \triangleq \frac{1}{2}(\mathbf{A}-\mathbf{A}^H)$ are the real and imaginary parts of the matrix $\mathbf{A} = \Re(\mathbf{A})+j\Im(\mathbf{A})$. $\mathtt{P}$ represents the probability. $f_X(\cdot)$ is the probability density function (PDF) of a random variable X. $\mathtt{E}[X]$ is the expected value of X. $\mathbf{h} \simeq \mathtt{CN}(\mathbf{K})$ means that $\mathbf{h}$ is a circulary-symmetric complex Gaussian random vector with covariance matrix $\mathbf{K}$. $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ is the all-zero matrix. For any sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A}-\mathcal{B}$ is the set of elements in $\mathcal{A}$, but not in $\mathcal{B}$. $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$. $\emptyset$ is the empty set. For any logical statement $\mathtt{ST}$, $\mathbf{1}(\mathtt{ST}) = 1$ if $\mathtt{ST}$ is true, and $\mathbf{1}(\mathtt{ST}) = 0$, otherwise. Finally, $\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{u^2}{2}}\mathrm{d}u$, $x \in \mathbb{R}$ is the Gaussian tail function, $\log(\cdot)$ is the natural logarithm, $\log_2(\cdot)$ is the logarithm to base 2, and $\Gamma(\cdot)$ is the gamma function.

II. Preliminaries

A. System Model

Consider a $t \times 1$ MISO system. Denote the channel from transmitter antenna i to the receiver antenna by h_i, and let $\mathbf{h} = [\begin{array}{ccc} h_1 & \cdots & h_t \end{array}]^T \in \mathbb{C}^{t\times 1}$ represent the channel state. Throughout the paper, we assume that $\mathbf{h} \simeq \mathtt{CN}(\mathbf{I})$. We also assume that the fading is sufficiently slow so that the receiver can estimate the channel state and feed back its quantized version to the transmitter.

Let $s \in \mathbb{C}$ denote the information-bearing symbol that we wish to transmit. Unless otherwise specified, we assume that $s \simeq \mathtt{CN}(1)$. In other words, we assume that s is a circulary-symmetric complex Gaussian symbol with unit energy. For a given channel state $\mathbf{h}$, suppose that the transmitter sends $s\mathbf{x}^\dagger\sqrt{P}$ over its t antennas, where $\mathbf{x} \in \chi$ is a beamforming vector and

$$\chi \triangleq \{\mathbf{x} \in \mathbb{C}^{t\times 1} : \|\mathbf{x}\| = 1\} \tag{1}$$

is the set of all feasible beamforming vectors. The channel input-output relationship with such a transmission strategy can be expressed as

$$y = s\langle\mathbf{h},\mathbf{x}\rangle\sqrt{P} + \eta, \tag{2}$$

where y is the received signal, $\eta \in \mathtt{CN}(1)$ is the noise at the receiver, and P is the ratio of the short term power constraint of the transmitter to the noise power at the receiver. Note that since the noise power is normalized to unity, P is also equal to the short-term power constraint of the transmitter.

For a fixed channel state $\mathbf{h}$ and a fixed beamforming vector $\mathbf{x}$, the mutual information of the MISO channel in (2) is $\log_2(1+|\langle \mathbf{x},\mathbf{h}\rangle|^2 P)$ bits per channel use. For a given target data transmission rate ρ, we say that an outage event occurs if $\log_2(1+|\langle \mathbf{x},\mathbf{h}\rangle|^2 P) < \rho$.

When $\mathbf{h}$ is random, we can choose a different beamforming vector for different $\mathbf{h}$. In this case, we define the outage probability as the fraction of channel realizations for which outage events occur. The clear goal is to minimize this outage probability. Formally, consider an arbitrary mapping $\mathtt{m}:\mathbb{C}^t \to \chi$. Then, the outage probability with mapping $\mathtt{m}$ can be expressed as

$$\mathtt{OUT}(\mathtt{m}) \triangleq \mathtt{P}(|\langle \mathtt{m}(\mathbf{h}),\mathbf{h}\rangle|^2 < \alpha), \tag{3}$$

where

$$\alpha \triangleq \frac{2^\rho - 1}{P}. \tag{4}$$

B. Full-CSIT and No-CSIT Systems

At one extreme case, the transmitter may know $\mathbf{h}$ perfectly, in which case we say that we have a "full-CSIT system." In such a scenario, we can choose an optimal beamforming vector, say $\mathtt{Full}(\mathbf{h})$ for a given $\mathbf{h}$. We have $|\langle \mathtt{Full}(\mathbf{h}),\mathbf{h}\rangle| \leq \|\mathbf{h}\|$, and the upper bound is achievable by choosing, for example,

$$\mathtt{Full}(\mathbf{h}) = \frac{\mathbf{h}^*}{\|\mathbf{h}\|}. \tag{5}$$

This gives us

$$\mathtt{OUT}(\mathtt{Full}) = \mathtt{P}(\|\mathbf{h}\|^2 < \alpha). \tag{6}$$

At the other extreme, the transmitter may not know $\mathbf{h}$ at all, in which case we say that we have a no-CSIT system or an open loop system. In this case, the transmitter uses a single beamforming vector for all the channel states. Therefore, a no-CSIT system can be described by the mapping $\mathtt{open}:\mathbb{C}^t \to \{\mathbf{o}\}$ for some $\mathbf{o} \in \chi$. We assume that such an open loop system is optimally designed in the sense that

$$\mathbf{o} = \arg\min_{\mathbf{x}\in\chi} \mathtt{P}(|\langle \mathbf{x},\mathbf{h}\rangle|^2 < \alpha) \tag{7}$$

with ties broken arbitrarily. For any $\mathbf{x} \in \chi$, we have $\langle \mathbf{x},\mathbf{h}\rangle \sim \mathtt{CN}(1)$, which implies $\mathtt{P}(|\langle \mathbf{x},\mathbf{h}\rangle|^2 < \alpha) = 1 - e^{-\alpha}$. In other words, the probability in the minimization in (7) does not depend on $\mathbf{x}$. Without loss of optimality, we may therefore set $\mathbf{o} = [1\ 0\ \cdots\ 0]^T$. This gives us $\mathtt{OUT}(\mathtt{open}) = \mathtt{P}(|h_1|^2 < \alpha) = 1 - e^{-\alpha}$.

We now consider the case where the transmitter has partial CSI via feedback from the receiver. Using a source coding formulation, such a partial CSI system can be described by a channel quantizer as we explain in what follows.

C. The Channel Quantizer

Let $\mathcal{I} \in \{\{0\},\{0,1\},\{0,1,2\},\ldots,\mathbb{N}\}$ be a possibly infinite index set whose elements are either the first $|\mathcal{I}|$ natural numbers or all the natural numbers. We use the notations $\{a_n\}_\mathcal{I}$ and $\{a_n : n \in \mathcal{I}\}$ interchangeably to represent a set whose elements are the real numbers a_n, $n \in \mathcal{I}$. Similar definitions hold for sets of vectors, collection of sets, etc.

For a given index set $\mathcal{I}$, let $\{\mathbf{x}_n\}_\mathcal{I}$ be a set of quantized beamforming vectors with $\{\mathbf{x}_n\}_\mathcal{I} \subset \chi$. Also, let $\{\mathcal{E}_n\}_\mathcal{I}$ with $\mathcal{E}_n \subset \mathbb{C}^t, \forall n \in \mathbb{N}$ be a collection of mutually disjoint measurable subsets of $\mathbb{C}^t$ with $\bigcup_{n\in\mathcal{I}} \mathcal{E}_n = \mathbb{C}^t$. Finally, let $\{\mathtt{b}_n\}_\mathcal{I}$ be a collection of feedback binary codewords with $\{\mathtt{b}_n\}_\mathcal{I} \subset \{0,1\}^\star$, where $\{0,1\}^\star \triangleq \{\epsilon,\mathtt{0},\mathtt{1},\mathtt{00},\mathtt{01},\ldots\}$ is the set of all binary codewords including the empty codeword ϵ. We assume $\mathtt{b}_m \neq \mathtt{b}_n$ whenever $m \neq n$. We call the collection of triples

$$\mathtt{q} := \{\mathbf{x}_n,\, \mathcal{E}_n,\, \mathtt{b}_n\}_\mathcal{I} \tag{8}$$

a quantizer $\mathtt{q}$ for the beamforming strategy. We call $\mathtt{q}$ an infinite-level quantizer if $\mathcal{I}$ is an infinite set. Otherwise, we call $\mathtt{q}$ an $|\mathcal{I}|$-level quantizer.

The quantizer definition in (8) immediately induces a feedback transmission scheme that operates in the following manner: For a fixed channel state $\mathbf{h}$, the receiver feeds back the binary codeword $\mathtt{b}_n$, where the index n here satisfies $\mathbf{h} \in \mathcal{E}_n$. Such an index n always exists and is unique as $\mathcal{E}_n$, $n \in \mathbb{N}$ is a disjoint covering of $\mathbb{C}^t$. The transmitter recovers the index n and uses the corresponding beamforming vector $\mathbf{x}_n$. The recovery of n by the transmitter is always possible since $\mathtt{b}_n$s are distinct. We write $\mathtt{q}(\mathbf{h}) = \mathbf{x}_n$ whenever $\mathbf{h} \in \mathcal{E}_n$ to emphasize the quantization operation. We call the set $\{\mathbf{x}_n\}_\mathcal{I}$ the quantizer (or beamforming) codebook.

For any $\mathtt{b} \in \{0,1\}^\star$, let $\mathtt{L}(\mathtt{b})$ denote the "length" of $\mathtt{b}$. For example, $\mathtt{L}(\epsilon) = 0$, $\mathtt{L}(\mathtt{01}) = 2$. A quantizer $\mathtt{q}$ is called an FLQ if $\mathtt{L}(\mathtt{b}_m) = \mathtt{L}(\mathtt{b}_n)$, $\forall m,n \in \mathcal{I}$. Otherwise, we call $\mathtt{q}$ a VLQ.

A quantizer $\mathtt{q}$ is thus a mapping $\mathbb{C}^t \to \{\mathbf{x}_n\}_\mathcal{I}$ supplied with a feedback binary codeword $\mathtt{b}_n$ for each $\mathbf{x}_n$. Treated solely as a mapping, it is a special case of the mapping $\mathtt{m} : \mathbb{C}^t \to \chi$ discussed in the previous section with the requirement of a countable image $\{\mathbf{x}_n\}_\mathcal{I}$. According to (3), we can therefore calculate the outage probability with $\mathtt{q}$ as $\mathtt{OUT}(\mathtt{q}) = \mathtt{P}(|\langle \mathtt{q}(\mathbf{h}),\mathbf{h}\rangle|^2 < \alpha)$, which does not depend on $\mathtt{b}_n$. As a result, we do not specify/mention $\mathtt{b}_n$ when we would like to talk only about the outage performance of a quantizer and write $\{\mathbf{x}_n,\mathcal{E}_n,\cdot\}_\mathcal{I}$ instead. The binary codewords $\mathtt{b}_n$ however determine the rate $\mathtt{R}(\mathtt{q})$ of the quantizer $\mathtt{q}$ by the formula

$$\mathtt{R}(\mathtt{q}) \triangleq \sum_{n\in\mathcal{I}} \mathtt{P}(\mathbf{h} \in \mathcal{E}_n)\mathtt{L}(\mathtt{b}_n). \tag{9}$$

We measure the quality of the quantizer by the outage probability it provides. We can also define the "binary" distortion measure

$$d(\mathbf{h},\mathbf{x}) = \mathbf{1}(|\langle \mathbf{x},\mathbf{h}\rangle|^2 < \alpha \text{ and } \|\mathbf{h}\|^2 \geq \alpha) \tag{10}$$

$$= \mathbf{1}(|\langle \mathbf{x},\mathbf{h}\rangle|^2 < \alpha) - \mathbf{1}(\|\mathbf{h}\|^2 < \alpha) \in \{0,1\} \tag{11}$$

that measures the quality of reproduction of the channel sample $\mathbf{h}$ by $\mathbf{x}$. For a given quantizer $\mathtt{q}$, the expected distortion

$\mathtt{E}[d(\mathbf{h},\mathtt{q}(\mathbf{h}))]$ is nothing but the quantity $\mathtt{OUT}(\mathtt{q})-\mathtt{OUT}(\mathtt{Full})$. Therefore, minimizing the expected distortion with $\mathtt{q}$ is equivalent to minimizing the outage probability with $\mathtt{q}$.

It is well-known that the outage performance of any finite-level quantizer is strictly worse than the full-CSIT performance $\mathtt{OUT}(\mathtt{Full})$ [28]. In other words, any finite-level quantizer has non-zero distortion. Hence, if we would like to achieve $\mathtt{OUT}(\mathtt{Full})$, we need to use an infinite-level quantizer with encoding regions $\{\mathcal{E}_n\}_{\mathbb{N}}$ that satisfy $\mathtt{P}(\mathbf{h}\in\mathcal{E}_n)>0$ for infinitely many n (As otherwise, we may ignore the $\mathcal{E}_n$ with $\mathtt{P}(\mathbf{h}\in\mathcal{E}_n)=0$ and obtain a finite-level quantizer that achieves $\mathtt{OUT}(\mathtt{Full})$, which is a contradiction.) Unfortunately, the rate of an FLQ with such encoding regions cannot be finite. Therefore, FLQs cannot achieve the full-CSIT performance with finite rate, and we have to consider VLQs for this purpose. Fortunately, the rate of an infinite-level VLQ may be finite even if all its encoding regions have non-zero probability.

We show that there is indeed a VLQ that achieves $\mathtt{OUT}(\mathtt{Full})$ with finite rate. We provide an explicit construction of such a VLQ. We also estimate the minimum rate at which $\mathtt{OUT}(\mathtt{Full})$ is achievable as a function of α, or equivalently, as a function of P. We start with the design of the encoding regions for a given codebook.

III. A New Encoding Rule

In this section, we introduce a new encoding rule for VLQs. In order to motivate the new encoding rule, we first review the standard encoding rule used for FLQs and discuss why it will not work in the case of VLQs.

A. The Standard Encoding Rule

Let $\mathcal{B}=\{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\}$ be a finite-cardinality beamforming codebook. In the case of FLQs, the standard encoding rule (see e.g. [6], [28]) is to choose the beamforming vector in $\mathcal{B}$ that is "closest" to $\mathbf{h}$ with respect to the absolute inner product. In other words, the standard approach is to work with the quantizer

$$\bar{\mathtt{q}}_{\mathcal{B}}(\mathbf{h})\triangleq\arg\max_{\mathbf{x}\in\mathcal{B}}|\langle\mathbf{x},\mathbf{h}\rangle|,\tag{12}$$

with ties broken arbitrarily. It can be shown that $\bar{\mathtt{q}}_{\mathcal{B}}$ is an optimal quantizer for the codebook $\mathcal{B}$ in the sense that for any quantizer $\mathtt{q}_{\mathcal{B}}:\mathbb{C}^t\to\mathcal{B}$, we have $\mathtt{OUT}(\bar{\mathtt{q}}_{\mathcal{B}})\leq\mathtt{OUT}(\mathtt{q}_{\mathcal{B}})$.

One way to design a VLQ might be to keep the standard encoding rule but instead use a variable-length code instead of a fixed-length code. There are two problems with this approach. The first problem, which is of a rather technical nature, is that the standard encoding rule is not well-defined for infinite-level quantizers as a maximizer may not exist for countably infinite codebooks. The second and more important issue is that even for a finite cardinality codebook, this rule is quite ill-suited for variable-length quantization as we shall discuss in the following.

It is well-known that the beamforming vectors in a well-designed codebook should be "evenly distributed" on χ (A formal treatment of this argument gives rise to e.g. Grassmannian codebooks [2]). For such a well-designed codebook $\mathcal{B}=\{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\}$ and an index $i\in\{0,\ldots,N-1\}$, the standard encoder picks $\mathbf{x}_i$ if $\forall n\in\{0,\ldots,N-1\}$, $|\langle\mathbf{x}_i,\mathbf{h}\rangle|\geq|\langle\mathbf{x}_n,\mathbf{h}\rangle|$. Due to the even distribution of codevectors, this results in quantization cells with roughly equal probability $\frac{1}{N}$. In such a scenario, it can be shown that even the best variable-length code results in a VLQ rate of $\log_2 N$ (up to an additive constant). Hence, VLQs designed via the standard encoding rule cannot achieve the full-CSIT performance with finite rate since we need $N\to\infty$ (In fact, we can already design a rate-$\lceil\log_2 N\rceil$ FLQ that is optimal for $\mathcal{B}$; a VLQ with almost the same rate is superfluous). We thus first introduce an alternate encoding strategy.

B. The New Encoding Rule

For any given beamforming vector $\mathbf{x}\in\chi$, let

$$\mathcal{O}_{\mathbf{x}}=\{\mathbf{h}\in\mathbb{C}^t:|\langle\mathbf{x},\mathbf{h}\rangle|^2<\alpha\}\tag{13}$$

denote the channel states for which using $\mathbf{x}$ results in outage. Also, let $\mathcal{O}_{\mathbf{x}}^c$ denote the complement of $\mathcal{O}_{\mathbf{x}}$. The simple but key observation is that the standard encoder is "excessively precise" as it always picks the (intuitively best) beamforming vector in $\{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\}$ that is closest to $\mathbf{h}$. In fact, without loss of optimality in terms of the outage probability, for any $j\in\{0,\ldots,N-1\}$, the transmitter can use $\mathbf{x}_j$ whenever using $\mathbf{x}_j$ does not result in outage (i.e. whenever $\mathbf{h}\in\mathcal{O}_{\mathbf{x}_j}^c$). It can also use $\mathbf{x}_j$ whenever all the beamforming vectors in the codebook result in outage (i.e. whenever $\mathbf{h}\in\bigcap_{n=0}^{N-1}\mathcal{O}_{\mathbf{x}_n}$), as using any other beamforming vector in $\{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\}$ would result in an inevitable outage anyway. In other words, for the set $\mathcal{O}_{\mathbf{x}_j}^c\cup\bigcap_{n=0}^{N-1}\mathcal{O}_{\mathbf{x}_n}$ of source samples, choosing $\mathbf{x}_j$ instead of the beamforming vector that is closest to $\mathbf{h}$ will not change the distortion. We exploit this property of the outage probability performance measure to design a new encoding strategy that yields low rates without sacrificing performance.

Formally, for a given arbitrary beamforming codebook $\{\mathbf{x}_n\}_{\mathcal{I}}$, we set

$$\mathcal{E}_0^\star\triangleq\mathcal{O}_{\mathbf{x}_0}^c\cup\bigcap_{n\in\mathcal{I}}\mathcal{O}_{\mathbf{x}_n},\tag{14}$$

and use $\mathbf{x}_0$ as the beamforming vector whenever $\mathbf{h}\in\mathcal{E}_0^\star$. We have now allocated the part $\mathcal{E}_0^\star$ of the channel state space $\mathbb{C}^t$. In general, whenever $|\mathcal{I}|\geq 2$, for any $n\in\mathcal{I}-\{0\}$, we set

$$\mathcal{E}_n^\star=\mathcal{O}_{\mathbf{x}_n}^c\cap\bigcap_{k=0}^{n-1}\mathcal{O}_{\mathbf{x}_k},\tag{15}$$

and use $\mathbf{x}_n$ whenever $\mathbf{h}\in\mathcal{E}_n^\star$. For any $n\in\mathcal{I}-\{0\}$, by definition, $\mathcal{E}_n^\star$ consists of channel states for which using the beamforming vector $\mathbf{x}_n$ does not result in outage while using any of the preceding beamforming vectors $\mathbf{x}_0,\ldots,\mathbf{x}_{n-1}$ results in outage.

It follows immediately from the definitions that $\{\mathcal{E}_n^\star\}_{\mathcal{I}}$ is a disjoint collection of measurable sets that cover $\mathbb{C}^t$. We may therefore define the (possibly infinite-level) quantizer $\{\mathbf{x}_n,\mathcal{E}_n^\star,\cdot\}_{\mathcal{I}}$.

An alternate more natural definition can be given as follows: The quantizer $\{\mathbf{x}_n,\mathcal{E}_n^\star,\cdot\}_{\mathcal{I}}$ selects the beamforming vector

$$\arg\min_{\mathbf{x}\in\{\mathbf{x}_n\}_{\mathcal{I}}}d(\mathbf{h},\mathbf{x})=\arg\min_{\mathbf{x}\in\{\mathbf{x}_n\}_{\mathcal{I}}}\mathbf{1}(|\langle\mathbf{x},\mathbf{h}\rangle|^2<\alpha),\tag{16}$$

with ties broken in favor of the vector with the smallest index. In this form, the new encoding rule resembles the standard encoding rule we have discussed before, but there is a key difference: It is now specifically tailored for the outage probability measure, and is no longer a nearest-neighbor encoding rule in an inner-product-distance sense. However, it can still be considered to be a nearest-neighbor encoding rule with a "distance function" that can only take the values 0 and 1.

Let us now calculate the outage probability with the quantizer $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_\mathcal{I}$. Meanwhile, we show that it is also in fact an optimal quantizer for the codebook $\{\mathbf{x}_n\}_\mathcal{I}$.

Proposition 1. *Let $\{\mathbf{x}_n\}_\mathcal{I}$ be a given codebook. For any quantizer $\mathtt{q} : \mathbb{C}^t \to \{\mathbf{x}_n\}_\mathcal{I}$, we have*

$$\mathtt{OUT}(\mathtt{q}) \geq \mathtt{P}\left(\mathbf{h} \in \bigcap_{i\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_i}\right). \tag{17}$$

Furthermore,

$$\mathtt{OUT}(\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_\mathcal{I}) = \mathtt{P}\left(\mathbf{h} \in \bigcap_{i\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_i}\right), \tag{18}$$

and therefore $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_\mathcal{I}$ is an optimal quantizer for $\{\mathbf{x}_n\}_\mathcal{I}$.

Proof. For any quantizer $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \cdot\}_\mathcal{I}$, the event $\mathbf{h} \in \bigcap_{n\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_n}$ results in outage regardless of how $\{\mathcal{E}_n\}_\mathcal{I}$ is chosen. This proves the lower bound. As for the quantizer $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_\mathcal{I}$, by construction, an outage event happens if and only if $\mathbf{h} \in \mathcal{E}_0^\star = \mathcal{O}_{\mathbf{x}_0}^c \cup \bigcap_{n\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_n}$, for which the transmitter uses the beamforming vector $\mathbf{x}_0$. Since $\mathbf{x}_0$ does not result in outage when $\mathbf{h} \in \mathcal{O}_{\mathbf{x}_0}^c$, we have the desired result. ■

The notion of "optimality for a given codebook" does not incorporate the rate of the quantizer. For example, both the standard encoder and our new encoding rule yield optimal quantizers for a given codebook, but which one has the lowest feedback rate? The main question here is to determine the optimal encoding structure in a rate-distortion (feedback rate-outage probability) sense. Formally, for any given target outage probability $p \in [\mathtt{OUT}(\mathtt{Full}), 1]$, let

$$\mathtt{r}^\star(p) = \inf_{\mathtt{q}} \{\mathtt{R}(\mathtt{q}) : \mathtt{OUT}(\mathtt{q}) \leq p\}, \tag{19}$$

denote the infimum of rates of all quantizers that can achieve an outage probability of at most p. Note that the quantity $\mathtt{r}^\star(p)$ depends on α (and P). A quantizer $\mathtt{q}$ is then called p-optimal (or simply optimal) if $\mathtt{OUT}(\mathtt{q}) = p$ and $\mathtt{R}(\mathtt{q}) = \mathtt{r}^\star(p)$. Existence of optimal quantizers is a non-trivial problem and will not be discussed here. We can however investigate the structure of these quantizers under the assumption that they exist.

We show in the following that if an $\mathtt{OUT}(\mathtt{Full})$-optimal quantizer exists, then there is another $\mathtt{OUT}(\mathtt{Full})$-optimal quantizer whose encoding regions are in the form given by (14) and (15). In other words, in the search for an $\mathtt{OUT}(\mathtt{Full})$-optimal quantizer, which is the main focus of this work, we may confine ourselves to the encoding strategy given by (14) and (15) without loss of generality or optimality. Meanwhile, we consider a slightly more general scenario, and find a similar "sufficient" encoding strategy for an arbitrary p-optimal quantizer, where $\mathtt{OUT}(\mathtt{Full}) < p \leq 1$. These optimality results will provide a theoretical basis for our "intituive choice" of the encoding rule. The uninterested reader may therefore skip to Section IV, as these optimality results will not play any role in the actual construction of quantizers.

C. Optimality of the New Encoding Rule

Let $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_\mathcal{I}$ be an arbitrary quantizer (not necessarily optimal for $\{\mathbf{x}_n\}_\mathcal{I}$) that satisfies $\mathtt{L}(\mathtt{b}_n) \leq \mathtt{L}(\mathtt{b}_{n+1}), \forall n, (n+1) \in \mathcal{I}$ without loss of generality. Note that an outage event with $\mathtt{q}$ happens if and only if $\mathbf{h}$ is a member of the set

$$\mathcal{O} \triangleq \bigcup_{n\in\mathcal{I}} (\mathcal{E}_n \cap \mathcal{O}_{\mathbf{x}_n}). \tag{20}$$

We thus have $\mathtt{OUT}(\mathtt{q}) = \mathtt{P}(\mathbf{h} \in \mathcal{O})$.

Let us now define the encoding regions

$$\mathcal{E}_0' = \mathcal{O}_{\mathbf{x}_0}^c \cup \mathcal{O}, \tag{21}$$

$$\mathcal{E}_n' = \mathcal{O}_{\mathbf{x}_n}^c \cap \bigcap_{k=0}^{n-1} \mathcal{O}_{\mathbf{x}_k} \cap \mathcal{O}^c,\ n \in \mathcal{I} - \{0\}. \tag{22}$$

It is straightforward to show that $\{\mathcal{E}_n'\}_\mathcal{I}$ is a disjoint covering of $\mathbb{C}^t$. Now, let

$$\mathtt{q}' = \{\mathbf{x}_n, \mathcal{E}_n', \mathtt{b}_n\}_\mathcal{I}. \tag{23}$$

We have the following result.

Proposition 2. *We have $\mathtt{OUT}(\mathtt{q}') \leq \mathtt{OUT}(\mathtt{q})$ and $\mathtt{R}(\mathtt{q}') \leq \mathtt{R}(\mathtt{q})$.*

Proof. See Appendix A. ■

For a given $p \in [\mathtt{OUT}(\mathtt{Full}), 1]$, if a p-optimal quantizer exists, then according to Proposition 2, we can find another p-optimal quantizer whose encoding regions are given by (21) and (22). Hence, in the quest of finding optimal quantizers, it is sufficient to consider quantizers with encoding regions of the form given by (21) and (22) (for some $\mathcal{O}$).

Also, in such a quest, a natural approach might be to consider quantizers that are optimal for their codebooks. Then, if $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_\mathcal{I}$ is an optimal quantizer for $\{\mathbf{x}_n\}_\mathcal{I}$, we have $\mathcal{O} = \bigcap_{n\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_n}$ up to a set of (probability) measure zero. Substituting $\mathcal{O} = \bigcap_{n\in\mathcal{I}} \mathcal{O}_{\mathbf{x}_n}$ in (21) and (22), we obtain $\mathcal{E}_n' = \mathcal{E}_n^\star, \forall n \in \mathcal{I}$. Hence, quantizers that employ the encoding rule specified by (14) and (15) are in fact optimal among all quantizers that are optimal for their codebooks.

Moreover, for the class of quantizers that can achieve the full-CSIT performance, we need $\mathcal{O} = \{\mathbf{h} : \|\mathbf{h}\|^2 < \alpha\}$ up to a set of measure zero, regardless of whether or not the quantizer is optimal for its codebook. In this case, the encoding rules that correspond to $\{\mathcal{E}_n'\}_\mathcal{I}$ and $\{\mathcal{E}_n^\star\}_\mathcal{I}$ coincide. This verifies our claim that in the search for an $\mathtt{OUT}(\mathtt{Full})$-optimal quantizer, it is sufficient to consider the encoding strategy given by (14) and (15).

IV. Construction of VLQs

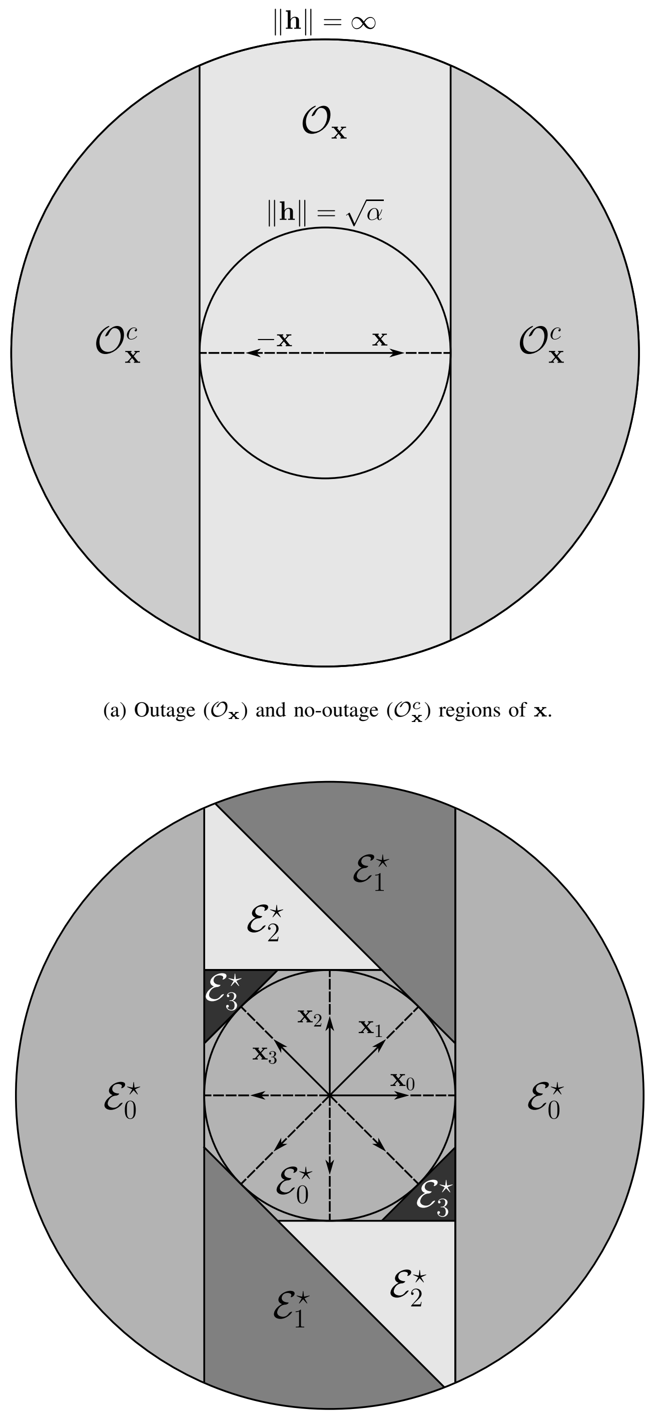
Having determined the optimal encoding rule, we now construct VLQs that can achieve $\texttt{OUT}(\texttt{Full})$ with low rates. Here, we argue that with an appropriate choice for $\{\mathbf{x}_n\}_{\mathbb{N}}$ and the feedback binary codewords $\{\mathtt{b}_n\}_{\mathbb{N}}$ the quantizer $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_{\mathbb{N}}$ can achieve $\texttt{OUT}(\texttt{Full})$ with finite rate. We first provide a graphical sketch that verifies our argument.

A. A Graphical Sketch

For a simpler illustration, we first consider a MISO system with two transmitter antennas and assume that the channel state $\mathbf{h}$ is a real-valued random vector in the two-dimensional space $\mathbb{R}^2$ (The PDF of this real-valued $\mathbf{h}$ is irrelevant for our discussion here; any PDF will "work" as long as it is bounded from above.). In such a scenario, the regions $\mathcal{O}_{\mathbf{x}} = \{\mathbf{h} : |\langle \mathbf{x}, \mathbf{h}\rangle| < \sqrt{\alpha}\}$ and $\mathcal{O}_{\mathbf{x}}^c$ are as shown in Fig. 1a for some $\alpha > 1$. In the figure, the entire space $\mathbb{R}^2$ is represented by the interior of the outer disk $\|\mathbf{h}\| \leq \infty$ that is bounded by the circle $\|\mathbf{h}\| = \infty$. The inner disk represents the set $\{\mathbf{h} : \|\mathbf{h}\|^2 \leq \alpha\}$. In fact, the probability that $\mathbf{h}$ remains in the interior of this disk is the full-CSIT outage probability $\texttt{OUT}(\texttt{Full})$ (evaluated with respect to the PDF of the real-valued $\mathbf{h}$). The beamforming vector $\mathbf{x}$ resides on the circle $\|\mathbf{h}\| = 1$ (not shown). The lighter shaded region in the middle is $\mathcal{O}_{\mathbf{x}}$ and the remaining two darker shaded regions constitute $\mathcal{O}_{\mathbf{x}}^c$. The regions $\mathcal{O}_{\mathbf{x}}$ and $\mathcal{O}_{\mathbf{x}}^c$ are separated by the two parallel lines $\{\mathbf{h} : \langle \mathbf{x}, \mathbf{h}\rangle = -\sqrt{\alpha}\}$ and $\{\mathbf{h} : \langle \mathbf{x}, \mathbf{h}\rangle = \sqrt{\alpha}\}$.

According to (14) and (15), the encoding regions given a quantizer codebook $\{\mathbf{x}_0, \ldots, \mathbf{x}_3\}$ will then be as shown as in Fig. 1b. In the figure, $\mathcal{E}_0^\star$ comprises of the interior of the hexagon formed at the center, and the two half planes on the left and right sides of the figure. By Proposition 1, the probability that $\mathbf{h}$ remains in the interior of the hexagon is the outage probability with the quantizer $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_{0\leq n\leq 3}$. It is greater than the full-CSIT outage probability as the hexagon cannot "completely cover" the inner circle. Now, as shown in Fig. 2, at Step 0, we start with the codebook $\{\mathbf{x}_0, \ldots, \mathbf{x}_3\}$ in Fig. 1b, and at Step ℓ, we add $2^{\ell+1}$ new beamforming vectors in between the ones we had in Step $\ell - 1$. Repeating this process indefinitely gives us an infinite codebook $\{\mathbf{x}_n\}_{\mathbb{N}}$ with a layered structure.

If we were to draw the codebook $\{\mathbf{x}_n\}_{\mathbb{N}}$ as we drew $\{\mathbf{x}_0, \ldots, \mathbf{x}_3\}$ we would observe that now $\bigcap_{n\in\mathbb{N}} \mathcal{O}_{\mathbf{x}}$ coincides with $\|\mathbf{h}\| < \alpha$ (up to a null set). This means that $\{\mathbf{x}_n, \mathcal{E}_n^\star, \cdot\}_{\mathbb{N}}$ can achieve $\texttt{OUT}(\texttt{Full})$. We now specify the feedback binary codewords for each quantization cell. Let $\mathtt{b}_0^\star = \epsilon$, where ϵ is the empty codeword, $\mathtt{b}_1^\star = \mathtt{0}$, $\mathtt{b}_2^\star = \mathtt{1}$, $\mathtt{b}_3^\star = \mathtt{00}$, $\mathtt{b}_4^\star = \mathtt{01}$, and sequentially so on for all the feedback binary codewords in $\{0,1\}^\star$. We have $\mathtt{L}(\mathtt{b}_n^\star) = \lfloor \log_2(n+1)\rfloor$, and we consider the rate of the quantizer $\{\mathbf{x}_n, \mathcal{E}_n^\star, \mathtt{b}_n^\star\}_{\mathbb{N}}$. From our figure for the encoding regions on $\{\mathbf{x}_n\}_{\mathbb{N}}$, we would also observe that the probabilities $\mathtt{P}(\mathbf{h} \in \mathcal{E}_n^\star)$ decay rather fast. In fact, for our two-dimensional space, it can be shown that they decay as $\frac{1}{n^2}$. This means that the quantization rate $\sum_{n\in\mathbb{N}} \mathtt{P}(\mathbf{h} \in \mathcal{E}_n^\star)\mathtt{L}(\mathtt{b}_n^\star) \sim \sum_{n\in\mathbb{N}} \lfloor \log_2(n+1)\rfloor \frac{1}{n^2}$ remains finite, and concludes our proof by figures for $\mathbb{R}^2$.



(a) Outage ($\mathcal{O}_{\mathbf{x}}$) and no-outage ($\mathcal{O}_{\mathbf{x}}^c$) regions of $\mathbf{x}$.

(b) The encoding regions of a 4-level quantizer.

Fig. 1: An illustration of the new encoding rule.

Consider now our actual problem that takes place in $\mathbb{C}^t$. Compared to the $\mathbb{R}^2$-scenario described above, the only difference is that we shall consider inner products in $\mathbb{C}^t$ instead of in $\mathbb{R}^2$. We thus intuitively expect that a quantizer with our new encoding rule and a $\mathbb{C}^t$-analogoue of the layered codebook in Fig. 2 to achieve the full-CSIT performance with finite rate. In the following, we formally verify this intuition. We first construct a codebook that resides in $\mathbb{C}^t$ and has the layered

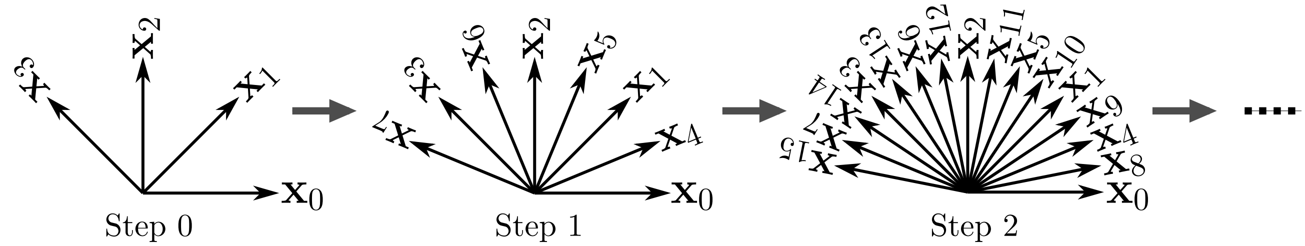


Fig. 2: A layered codebook structure.

structure in Fig. 2.

B. A Layered Codebook

For any $\ell \in \mathbb{N}$, let $\mathcal{S}_\ell = \{-1 + \frac{k}{2^{\ell+1}},\, k = 0, \ldots, 2^{\ell+2}\}$. For example, we have $\mathcal{S}_0 = \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$, and $\mathcal{S}_1 = \{-1, -\frac{3}{4}, -\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. For a given ℓ, we construct a codebook $\mathcal{Y}_\ell$ of beamforming vectors by setting

$$\mathcal{Y}_\ell = \left\{ \frac{\mathbf{y}}{\|\mathbf{y}\|} : \Re y_1,\, \Im y_1, \ldots, \Re y_t,\, \Im y_t \in \mathcal{S}_\ell,\, \|\mathbf{y}\| > 0 \right\}, \quad (24)$$

where $\Re y_i$ and $\Im y_i$ denote the real and imaginary parts of y_i, respectively. It is straightforward to verify that for any $\ell \in \mathbb{N}$,

$$\mathcal{Y}_\ell \subset \mathcal{Y}_{\ell+1}, \quad (25)$$

and

$$|\mathcal{Y}_\ell| \leq 2^{2t(\ell+3)}. \quad (26)$$

The sequence of codebooks $\{\mathcal{Y}_\ell\}_{\ell \in \mathbb{N}}$ looks like the sequence of layered codebooks in Fig. 2, with the index "ℓ" representing the layer. For large ℓ, $\mathcal{S}_\ell$ provides a fine quantization of the interval $[-1, 1]$. The corresponding $\mathcal{Y}_\ell$ is roughly a product quantizer codebook, and thus provides an increasingly finer quantization of χ as ℓ increases. Let us now verify this claim. We shall find, for any channel direction $\overline{\mathbf{h}} \in \chi$, a beamforming vector $\mathbf{y} \in \mathcal{Y}_\ell$ such that $|\langle \mathbf{y}, \overline{\mathbf{h}} \rangle|^2 \to 1$ as $\ell \to \infty$. For this purpose, for any $\ell \in \mathbb{N}$ and $x \in [-1, 1]$, let

$$q_\ell(x) \triangleq \operatorname{sign}(x) \frac{1}{2^{\ell+1}} \left\lfloor |x| 2^{\ell+1} \right\rfloor. \quad (27)$$

We have $q_\ell(x) \in \mathcal{S}_\ell$. For a given $\overline{\mathbf{h}} = [\overline{h}_1 \cdots \overline{h}_t]^T \in \chi$, we construct the vector

$$Q_\ell(\overline{\mathbf{h}}) \triangleq \begin{bmatrix} q_\ell(\Re \overline{h}_1) + j q_\ell(\Im \overline{h}_1) \\ \vdots \\ q_\ell(\Re \overline{h}_t) + j q_\ell(\Im \overline{h}_t) \end{bmatrix} \quad (28)$$

by applying $q_\ell(\cdot)$ to the real and imaginary parts of the components of $\overline{\mathbf{h}}$. We show in Appendix B that for any $\ell \geq \lceil \log_2(2t) \rceil$,

$$0 < \|Q_\ell(\overline{\mathbf{h}})\| \leq 1, \quad (29)$$

and

$$|\langle Q_\ell(\overline{\mathbf{h}}), \overline{\mathbf{h}} \rangle|^2 > 1 - \frac{2t}{2^\ell}. \quad (30)$$

Therefore, normalizing $Q_\ell(\overline{\mathbf{h}})$ as

$$\overline{Q}_\ell(\overline{\mathbf{h}}) \triangleq \frac{Q_\ell(\overline{\mathbf{h}})}{\|Q_\ell(\overline{\mathbf{h}})\|}, \quad (31)$$

we have

$$\overline{Q}_\ell(\overline{\mathbf{h}}) \in \mathcal{Y}_\ell \quad (32)$$

by the definition of $\mathcal{Y}_\ell$ in (24), and

$$|\langle \overline{Q}_\ell(\overline{\mathbf{h}}), \overline{\mathbf{h}} \rangle|^2 > 1 - \frac{2t}{2^\ell}. \quad (33)$$

We have just proved the following result.

Proposition 3. *For any $\ell \geq \lceil \log_2(2t) \rceil$,*

$$\forall \overline{\mathbf{h}} \in \chi,\, \exists \mathbf{y} \in \mathcal{Y}_\ell,\, |\langle \mathbf{y}, \overline{\mathbf{h}} \rangle|^2 > 1 - \frac{2t}{2^\ell}. \quad (34)$$

In other words, $\mathcal{Y}_\ell$ can indeed provide an arbitrarily fine quantization of χ for large enough ℓ.

We now "glue" the codebook layers $\mathcal{Y}_\ell$, $\ell \in \mathbb{N}$ together (in the same manner as in Fig. 2) to come up with a single codebook with a layered structure. For this purpose, for any $i \in \{1, \ldots, t\}$, let $\mathbf{e}_i = [e_{i1} \cdots e_{it}]^\dagger$ with $e_{ii} = 1$, $e_{ij} = 0, \forall j \neq i$ denote the beamforming vector that selects the ith transmitter antenna. We have $\mathbf{e}_i \in \mathcal{Y}_0, \forall i \in \{1, \ldots, t\}$ by the construction of $\mathcal{Y}_0$. In particular, $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{Y}_0$, and we may therefore construct a quantizer codebook $\{\mathbf{y}_n\}_{\mathbb{N}}$ that satisfies

$$\mathbf{y}_0 = \mathbf{e}_1, \quad (35)$$

$$\mathbf{y}_1 = \mathbf{e}_2, \quad (36)$$

and

$$\forall \ell \in \mathbb{N},\, \bigcup_{n=0}^{|\mathcal{Y}_\ell|-1} \{\mathbf{y}_n\} = \mathcal{Y}_\ell. \quad (37)$$

In other words, the "first" two elements $\mathbf{y}_0$ and $\mathbf{y}_1$ of $\mathcal{Y}_0$ are $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively (In fact, instead of $\mathbf{e}_1$ and $\mathbf{e}_2$, any two linearly independent beamforming vectors will work fine too. The motivation for choosing $\mathbf{e}_1$ and $\mathbf{e}_2$ is nothing but to simplify the analysis and avoid unnecessary technicalities.). Also, for any $\ell \in \mathbb{N}$, the "first" $|\mathcal{Y}_\ell|$ elements $\mathbf{y}_0, \ldots, \mathbf{y}_{|\mathcal{Y}_\ell|-1}$ of $\{\mathbf{y}_n\}_{\mathbb{N}}$ form the set $\mathcal{Y}_\ell$. Such a set $\{\mathbf{y}_n\}_{\mathbb{N}}$ always exists since $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{Y}_0$, and by Proposition 3, $\mathcal{Y}_\ell \subset \mathcal{Y}_{\ell+1}$ with $|\mathcal{Y}_\ell| < \infty$. In fact, it is also straightforward to construct $\{\mathbf{y}_n\}_{\mathbb{N}}$ by setting $\mathbf{y}_0 = \mathbf{e}_1$, $\mathbf{y}_1 = \mathbf{e}_2$,

$$\{\mathbf{y}_2, \ldots, \mathbf{y}_{|\mathcal{Y}_0|-1}\} = \mathcal{Y}_0 - \{\mathbf{e}_1, \mathbf{e}_2\}, \quad (38)$$

(It does not matter which element of $\mathcal{Y}_0 - \{\mathbf{e}_1, \mathbf{e}_2\}$ is set to be e.g. $\mathbf{y}_2$ as long as the equality holds), and

$$\{\mathbf{y}_{|\mathcal{Y}_\ell|}, \ldots, \mathbf{y}_{|\mathcal{Y}_{\ell+1}|-1}\} = \mathcal{Y}_{\ell+1} - \mathcal{Y}_\ell,\ \ell \geq 1. \quad (39)$$

We thus have our easily constructable quantizer codebook $\{\mathbf{y}_n\}_\mathbb{N}$. We now show that this codebook can achieve the full-CSIT performance with finite-rate.

C. The Main Achievability Result

Let $\{\mathcal{F}_n\}_\mathbb{N}$ be defined as in (14) and (15) with respect to $\{\mathbf{y}_n\}_\mathbb{N}$. Also, let $\{\mathtt{b}_n^\star\}_\mathbb{N}$ be as defined in Section III-B. Consider the quantizer $\mathtt{q}^\star \triangleq \{\mathbf{y}_n, \mathcal{F}_n, \mathtt{b}_n^\star\}_\mathbb{N}$. In the following theorem, we analyze the outage probability and the rate of $\mathtt{q}^\star$.

Theorem 1. *For any P, we have*

$$\mathtt{OUT}(\mathtt{q}^\star) = \mathtt{OUT}(\mathtt{Full}), \quad (40)$$

and

$$\mathtt{R}(\mathtt{q}^\star) \leq \exp\left(-\frac{2^\rho - 1}{P}\right)\left[\frac{2^\rho - 1}{P} + \mathtt{C}_0\left(\frac{2^\rho - 1}{P}\right)^2 + \mathtt{C}_0\left(\frac{2^\rho - 1}{P}\right)^t\right], \quad (41)$$

where $\mathtt{C}_0$ is a constant that is independent of ρ and P.

Proof. See Appendix C. ∎

Theorem 1 shows that the explicitly constructable variable-length quantizer $\mathtt{q}^\star$ achieves the full-CSIT performance with finite rate. Moreover, according to the rate upper bound on $\mathtt{R}(\mathtt{q}^\star)$ in Theorem 1, for a fixed ρ, we have $\mathtt{R}(\mathtt{q}^\star) \to 0$ when $P \to 0$ or $P \to \infty$. Also, it is straightforward to show that for any P, the upper bound is in fact bounded from above by a constant that is independent of P. Hence, regarding $\mathtt{r}^\star(p)$, which is the minimum rate that guarantees an outage probability of p (as defined in (19)), we can conclude that for any $P \geq 0$, $\rho \geq 0$, and $p \in [\mathtt{OUT}(\mathtt{Full}), 1]$, $\mathtt{r}^\star(p) \leq \mathtt{C}_1$ for some absolute constant $\mathtt{C}_1$ that is independent of P and ρ.

A natural question is then to determine the minimum rate that guarantees the full-CSIT performance. We discuss this problem in the next section.

D. P-Asymptotically Optimal Quantizers

In the following, consider a fixed ρ. It is difficult to calculate the minimum rate that guarantees the full-CSIT performance. We can however determine how this minimum rate behaves as $P \to \infty$. Such a result is useful as one is usually interested in the medium-to-high P regime, where the outage probability is naturally low. First, we state the following result.

Theorem 2. *For any quantizer $\mathtt{q} : \mathbb{C}^t \to \chi$, we have $\mathtt{OUT}(\mathtt{q}) \geq \mathtt{OUT}(\mathtt{open}) - \mathtt{R}(\mathtt{q})$.*

Proof. See Appendix D. ∎

We now let $\mathtt{r}^\star(\mathtt{OUT}(\mathtt{Full}))$ denote the infimum of the rates of those quantizers that achieve the full-CSIT performance, as defined in (19). We can then obtain the following result as a corollary to Theorems 1 and 2.

Corollary 1. *We have* $\lim_{P\to\infty}\left(\mathtt{r}^\star(\mathtt{OUT}(\mathtt{Full}))/\frac{2^\rho-1}{P}\right) = 1$.

Proof. See Appendix E. ∎

By Corollary 1, we thus conclude that the necessary and sufficient feedback rate to achieve the full-CSIT performance is $\sim \frac{2^\rho-1}{P}$. According to Theorem 1, the quantizer $\mathtt{q}^\star$ is thus a "P-asymptotically optimal" quantizer.

V. Implementation Issues

In the previous section, we have provided an explicit construction of a VLQ (namely $\mathtt{q}^\star$) that can achieve the full-CSIT performance with finite rate. The purpose of this section is to discuss and resolve some certain challenges that one may face in the process of an implementation of $\mathtt{q}^\star$. One can obviously come up with an arbitrarily long list of "practical implementation issues," and we shall neither attempt to address nor claim to solve all these issues in an exhaustive manner. Still, a certain aspect of $\mathtt{q}^\star$, namely the fact that it has an infinite codebook, immediately stands out as arguably the most problematic and critical for implementation purposes. We first discuss the practical challenges of implementing infinite codebooks and propose methods to resolve these challenges. Finally, we consider the practicalities of $\mathtt{q}^\star$ with non-prefix-free codes and offer prefix-free coded quantization as an alternative.

A. "Fast" Encoding and Decoding without Codebook Storage

A naive implementation of the quantizer $\mathtt{q}^\star = \{\mathbf{y}_n, \mathcal{F}_n, \mathtt{b}_n^\star\}_\mathbb{N}$, at least the way it is mathematically defined, requires both the transmitter and the receiver to store an infinite codebook. Moreover, even under the assumption that we can store such a codebook, the encoding process itself may require an arbitrarily large number of arithmetic operations. For example, suppose that the channel state $\mathbf{h}$ satisfies $\|\mathbf{h}\|^2 > \alpha$. In this case, we know that there is at least one beamforming vector in $\{\mathbf{y}_n\}_\mathbb{N}$ that can avoid outage (This is because the codebook $\{\mathbf{y}_n\}_\mathbb{N}$ is dense in χ.). Then, in order to implement the quantizer $\mathtt{q}^\star$ as it is, we shall determine the outage-avoiding beamforming vector with the smallest index. One way to determine this "first" outage-avoiding beamforming vector is to sequentially calculate all the SNR values $|\langle\mathbf{y}_0, \mathbf{h}\rangle|^2$, $|\langle\mathbf{y}_1, \mathbf{h}\rangle|^2, \ldots$ until we reach the beamforming vector that provides an SNR of at least α. This strategy however requires an arbitrarily large number of SNR calculations as, depending on $\mathbf{h}$, the first outage-avoiding beamforming vector may have an arbitrarily large index.

The discussion above suggests that even for a specific subset $\{\mathbf{h} : \|\mathbf{h}\|^2 > \alpha\}$ of channel states, implementing $\mathtt{q}^\star$ in a computationally efficient manner is a difficult, non-trivial problem. Instead of working with the theoretically-optimal (in the sense of Proposition 2) quantizer $\mathtt{q}^\star$ itself, our idea is to modify $\mathtt{q}^\star$ to come up with a new (albeit suboptimal) quantizer that similarly achieves $\mathtt{OUT}(\mathtt{Full})$ with rate $\sim \frac{1}{P}$. The new quantizer will use the same codebook $\{\mathbf{y}_n\}_\mathbb{N}$ as $\mathtt{q}^\star$, but it will have different encoding regions and feedback binary codewords that allow "fast" encoding and decoding.

Here, "fast" refers to the fact that for both the encoder and the decoder of the quantizer, the number of arithmetic operations (on real numbers) per channel state is bounded from above by a constant that is independent of the channel state. Moreover, with the new quantizer, neither the receiver nor the transmitter will have to store the infinite codebook $\{\mathbf{y}_n\}_{\mathbb{N}}$, even though in principle, the system will still be operating with $\{\mathbf{y}_n\}_{\mathbb{N}}$.

Note that we shall still retain the assumption that we can perform arbitrary precision arithmetic and store a finite number of real numbers. A finite-precision model brings in many issues (For example, the transmitter will not be able to use a Gaussian codebook for data transmission, and thus the mutual information and the outage probability expressions will be different. The receiver will not be able to store the channel state $\mathbf{h}$ perfectly; we shall therefore also take into account channel estimation errors, and design an alternate decoding method for the data symbols, etc.) that are irrelevant to the channel quantization process, and is therefore well-beyond the scope of this paper. Still, the methods that will be presented in this section can be applied to the finite precision case.

1) Construction of Fast VLQs for Beamforming: For the reader's convenience, we recall the auxiliary product-like quantizer $\overline{Q}_\ell(\overline{\mathbf{h}})$ discussed in Section IV-B. For a given channel direction $\overline{\mathbf{h}}$, it is the normalization $\overline{Q}_\ell(\overline{\mathbf{h}}) = Q_\ell(\overline{\mathbf{h}})/\|Q_\ell(\overline{\mathbf{h}})\|$, where the ith component of $Q_\ell(\overline{\mathbf{h}})$ is $q_\ell(\Re\overline{h}_i) + jq_\ell(\Im\overline{h}_i)$, and $q_\ell(x) = \mathrm{sign}(x)2^{-(\ell+1)}\left\lfloor |x|2^{\ell+1}\right\rfloor$.

We now need the following result.

Proposition 4. *For any $\|\mathbf{h}\|^2 > \alpha$, let*

$$\ell^\star \triangleq \max\left\{\lceil\log_2(4t)\rceil, \left\lceil -\log_2\left(\frac{1}{4t}\left(\frac{\|\mathbf{h}\|^2}{\alpha}-1\right)\right)\right\rceil\right\}. \quad (42)$$

Then, $|\langle\overline{Q}_{\ell^\star}(\overline{\mathbf{h}}), \mathbf{h}\rangle|^2 \geq \alpha$.

Proof. See Appendix F. ∎

In other words, for any channel state $\mathbf{h}$ with $\|\mathbf{h}\|^2 > \alpha$, the beamforming vector $\overline{Q}_{\ell^\star}(\overline{\mathbf{h}}) \in \mathcal{Y}_{\ell^\star}$ does not result in outage.

The proposition has a very simple interpretation: If the channel magnitude $\|\mathbf{h}\|$ is large enough, we can avoid outage by considering only the beamforming vectors in the low resolution layer $\ell^\star = \lceil\log_2(4t)\rceil$. On the other hand, if $\|\mathbf{h}\|^2$ is "close" to α, we need to consider higher resolution layers. The proposition gives us an estimate of the layer (via $\ell^\star$) where we can surely find a beamforming vector that avoids outage.

The simple product-like structure of $\overline{Q}_{\ell^\star}(\cdot)$ makes the tasks of finding this no-outage beamforming vector (that resides at layer $\ell^\star$) and communicating it to the transmitter very easy. Indeed, for a given $x \in [-1,1]$, and any $\ell \in \mathbb{N}$, we can easily calculate $q_\ell(x)$ by taking the most significant $\ell+2$ bits $(b_1.b_2b_3\cdots b_{\ell+1})_2$ of the binary representation $(b_1.b_2b_3\cdots)_2$ of $|x|$, while preserving the sign of x. For example, we have $q_1(\pm(0.101)_2) = \pm(0.10)_2$.

It follows that $Q_{\ell^\star}(\overline{\mathbf{h}})$ (and $\overline{Q}_{\ell^\star}(\overline{\mathbf{h}})$) can be uniquely represented by $2t(\ell^\star+3)$ bits since $\overline{\mathbf{h}}$ has a total of $2t$ real and complex dimensions. Assuming that $Q_{\ell^\star}(\overline{\mathbf{h}})$ is stored digitally, we communicate it to the transmitter uncoded. In other words, we represent $Q_{\ell^\star}(\overline{\mathbf{h}})$ by "itself;" a complicated bit assignment mapping is not necessary. The transmitter then normalizes $Q_{\ell^\star}(\overline{\mathbf{h}})$ to recover the no-outage beamforming vector $\overline{Q}_{\ell^\star}(\overline{\mathbf{h}})$.

Algorithm 1 A fast encoder for the layered codebook

if $|h_1|^2 \geq \alpha$ or $\|\mathbf{h}\|^2 \leq \alpha$ **then**
  Feed back the empty codeword ϵ.
else if $|h_2|^2 \geq \alpha$ **then**
  Feed back `0`.
else
  Calculate $\ell^\star$ as shown in (42).
  Feed back $Q_{\ell^\star}(\overline{\mathbf{h}})$.
end if

Algorithm 2 A fast decoder for the layered codebook

$\mathtt{b} \leftarrow$ Received feedback binary codeword.
if $\mathtt{b} = \epsilon$ **then**
  Use $\mathbf{e}_1$.
else if $\mathtt{b} = \mathtt{0}$ **then**
  Use $\mathbf{e}_2$.
else
  $\ell^\star \leftarrow \frac{\mathtt{L(b)}}{2t} - 3$, where $\mathtt{L(b)}$ is the length of $\mathtt{b}$.
  Reconstruct $Q_{\ell^\star}(\overline{\mathbf{h}})$.
  Use $\overline{Q}_{\ell^\star}(\overline{\mathbf{h}})$.
end if

These ideas lead to the quantizer encoding and decoding algorithms as shown in Algorithms 1 and 2, respectively. We call the corresponding quantizer $\mathtt{q}_{\mathtt{F}}^\star$. We analyze the performance of $\mathtt{q}_{\mathtt{F}}^\star$ in the following theorem.

Theorem 3. *For any $P > 0$, we have*

$$\mathtt{OUT}(\mathtt{q}_{\mathtt{F}}^\star) = \mathtt{OUT}(\mathtt{Full}), \quad (43)$$

and

$$\mathtt{R}(\mathtt{q}_{\mathtt{F}}^\star) \leq \exp\left(-\frac{2^\rho-1}{P}\right)\left[\frac{2^\rho-1}{P} + \mathtt{C}_2\left(\frac{2^\rho-1}{P}\right)^2 + \mathtt{C}_2\left(\frac{2^\rho-1}{P}\right)^t\right], \quad (44)$$

where $\mathtt{C}_2$ is a constant that is independent of ρ and P.

Proof. See Appendix G. ∎

Similarly to the quantizer $\mathtt{q}^\star$ in Section IV, the quantizer $\mathtt{q}_{\mathtt{F}}^\star$ can achieve the full-CSIT performance with the asymptotically-optimal rate $\sim \frac{2^\rho-1}{P}$. In addition, $\mathtt{q}_{\mathtt{F}}^\star$ has the advantage of fast encoders and decoders that do not have to store the quantizer codebook.

2) Structural Comparison of the Fast and the Optimal: It is instructive to draw analogies between the design principles and structures of $\mathtt{q}^\star$ and $\mathtt{q}_{\mathtt{F}}^\star$. In the optimal quantizer $\mathtt{q}^\star$, we use the beamforming vector $\mathbf{y}_n$ if and only if it yields zero distortion and all of the previous options $\mathbf{y}_0,\ldots,\mathbf{y}_{n-1}$ yield non-zero distortion. Instead of this finest possible vector-level approach of $\mathtt{q}^\star$, we consider the coarser layer-level approach for $\mathtt{q}_{\mathtt{F}}^\star$: We calculate the lowest possible layer $\ell^\star$ where we can make sure we can find a no-outage beamforming vector for any channel direction. Once this layer is determined, we use a simple product quantizer to determine the no-outage beamforming vector. Therefore, despite the fact that $\mathtt{q}_{\mathtt{F}}^{\star}$ is "highly non-uniform" across different layers, it looks like a uniform nearest neighbor quantizer within each layer. The associated encoding rule can thus be considered as a hybrid of the standard and the new encoding rules discussed in Section III.

B. On Non-Prefix-Free Codes

Throughout the paper, we have allowed our quantizers to use non-prefix-free codes that include the empty codeword. Since the feedback binary codewords corresponding to different channel states are seperated in time, one can -at least in theory- "safely" use a non-prefix-free code, e.g. $\{\epsilon, \mathtt{0}, \mathtt{01}\}$. Such codes may however bring along some certain practical issues that we discuss below.

First, it is not immediately clear, at least in practice, whether or not we may employ the empty codeword "with 0 cost" as implied by the rate definition in (9). Perhaps, "the non-existence of information" may be regarded as an empty codeword of length 0. Such an argument makes perfect sense in interrupt-driven systems, such as computers, in which the non-existence of an external input (e.g. not pressing any of the keys on the keyboard) will not require any information transmission. Hence, if one treats the feedback information as a certain bit sequence that triggers an interrupt at the recipient of the feedback information (the transmitter), then having no feedback information, or equivalently an empty codeword makes sense.

Regardless of whether one allows the empty codeword or not, with a non-prefix-free code, the quantizer decoder may have difficulty in determining the length of the feedback codeword that it receives, and this may cause synchronization problems. For example, suppose that the quantizer decoder at the transmitter has received the feedback bits $\mathtt{001}$. It is not clear whether the decoder should wait for more bits, or the feedback message is complete and the intended feedback bits are in fact just $\mathtt{001}$. One way to resolve this issue might be to force the quantizer decoder to wait for a certain amount of time before declaring that the feedback message is complete.

An alternate way to resolve the above practical issues is to impose a prefix-free code for the quantizer binary codewords. Note that a prefix-free code cannot contain the empty codeword and it also resolves the synchronization problem discussed above. Our results can easily be extended to the case where the code is constrained to be prefix-free. In such a scenario, the conclusion that the full-CSIT performance is achievable with a finite rate still remains as it is. For a fixed ρ, the necessary and sufficient feedback rate that guarantees the full-CSIT performance becomes 1 bit per channel state for $P \to \infty$ and $P \to 0$.

Finally, another practical issue is that it may be hard to do resource allocation in the feedback link as the length of the feedback codewords can be any natural number: In most of the current wireless systems, the number of feedback bits available for each channel state is fixed. After we introduced the use and benefits of VLQs for beamforming, a solution to this practical resource allocation problem has been proposed in [35] in the context of the 802.11 framework. The idea of [35] is that since most wireless communication standards such as 802.11 support variable-length data packets, one may, at least in principle, consider a variable-length packet-based feedback scheme without leaving the confines of the existing standards. The details on how such a feedback scheme can actually be implemented is however well beyond the scope of this paper and can be found in [35].

VI. Extensions to Finite Symbol Alphabets

In the previous sections, we have designed outage-minimizing VLQs for a MISO system that employs Gaussian symbols for data transmission. In this section, we study the practically more relevant case of a discrete input distribution with finite support, such as a QAM or PSK constellation.

Consider the same MISO channel in (2). The difference is that we let $s \in \mathcal{S}$, where $\mathcal{S}$ is an arbitrary subset of $\mathbb{C}$ with finite cardinality (For example, $\mathcal{S} = \{-1, +1\}$ corresponds to a BPSK constellation.). As in [36], we assume that the probability density function of s is fixed and does not depend on the channel state $\mathbf{h}$. Also, without loss of generality, we assume that s is uniformly distributed on $\mathcal{S}$ (Our results can be extended to arbitrary probability distributions on $\mathcal{S}$.) with $\mathtt{E}[|s|^2] \leq 1$.

In this scenario, for a fixed channel state $\mathbf{h}$ and a fixed beamforming vector $\mathbf{x}$, the mutual information of the MISO channel in (2) is given by $\mathtt{C}(|\langle \mathbf{x}, \mathbf{h}\rangle|^2 P)$, where, for any $\mathtt{snr} > 0$, we let $\mathtt{C}(\mathtt{snr}) \triangleq I(s; s\sqrt{\mathtt{snr}} + n)$ denote the mutual information between s and $s\sqrt{\mathtt{snr}} + \eta$. For a given target data transmission rate ρ, an outage event occurs if $\mathtt{C}(|\langle \mathbf{x}, \mathbf{h}\rangle|^2 P) < \rho$. For a given mapping $\mathtt{m}$, we let $\mathtt{OUT}_{\mathcal{S}}(\mathtt{m}) \triangleq \mathtt{P}(\mathtt{C}(|\langle \mathtt{m}(\mathbf{h}), \mathbf{h}\rangle|^2 P) < \rho)$ denote the outage probability with $\mathtt{m}$.

In order to determine the minimum (full-CSIT) outage probability, or in general, in order to evaluate $\mathtt{OUT}_{\mathcal{S}}(\mathtt{m})$ for a given $\mathtt{m}$, we should be able to evaluate $\mathtt{C}(\mathtt{snr})$ for any given $\mathtt{snr} > 0$. Unfortunately, an explicit closed form formula for $\mathtt{C}(\mathtt{snr})$ is not available for an arbitrary alphabet $\mathcal{S}$. On the other hand, at least it is known (see e.g. [37]) that $\mathtt{C}(\mathtt{snr})$ is a monotonically increasing continuous function of $\mathtt{snr}$ with $\mathtt{C}(0) = 0$, $\mathtt{C}(\mathtt{snr}) < \log_2 |\mathcal{S}|, \forall \mathtt{snr}$, and $\lim_{\mathtt{snr}\to\infty} \mathtt{C}(\mathtt{snr}) = \log_2 |\mathcal{S}|$. In particular, the inequality implies that if $\rho \geq \log_2 |\mathcal{S}|$, we have $\mathtt{OUT}_{\mathcal{S}}(\mathtt{m}) = 1$ for any $\mathtt{m}$. From now on, we thus assume that $\rho < \log_2 |\mathcal{S}|$.

Since $\mathtt{C}(\mathtt{snr})$ is a monotonically increasing continuous function, its inverse function $\mathtt{C}^{-1}(\cdot)$ exists and we have, for any $\rho \in [0, \log_2 |\mathcal{S}|)$,

$$\mathtt{OUT}_{\mathcal{S}}(\mathtt{m}) = \mathtt{P}(|\langle \mathtt{m}(\mathbf{h}), \mathbf{h}\rangle|^2 P < \mathtt{C}^{-1}(\rho)) \tag{45}$$

$$= \mathtt{P}\left(|\langle \mathtt{m}(\mathbf{h}), \mathbf{h}\rangle|^2 < \beta\right), \tag{46}$$

where $\beta = \frac{\mathtt{C}^{-1}(\rho)}{P}$. The outage probability expression in (46) is in the exact same form as (3). Hence, all of the previous results that we have derived for Gaussian inputs can easily be extended to arbitrary inputs if the constant α (or $\frac{1}{P}$) that appears in the previous sections is replaced by β. In particular, the full-CSIT outage probability $\mathtt{P}(\|\mathbf{h}\|^2 < \beta)$

is achievable with a variable-length quantizer $\mathtt{q}_0$ with rate $\mathtt{R}(\mathtt{q}_0) \leq e^{-\beta}\left[\beta + \mathtt{C}_0\left(\beta^2+\beta^t\right)\right]$ (c.f. Theorem 1). In particular, $\mathtt{R}(\mathtt{q}_0) \sim \frac{\mathtt{C}^{-1}(\rho)}{P} \to 0$ as $P \to \infty$, and such a feedback rate is asymptotically the best possible (c.f. Corollary 1).

VII. Numerical Results

In this section, we present numerical evidence that verify our analytical results. For this purpose, we have simulated the quantizer $\mathtt{q}_{\mathtt{F}}^\star$ as it is defined via Algorithms 1 and 2 for $\rho = 1$ and different values of P. We have used the standard double precision arithmetic.

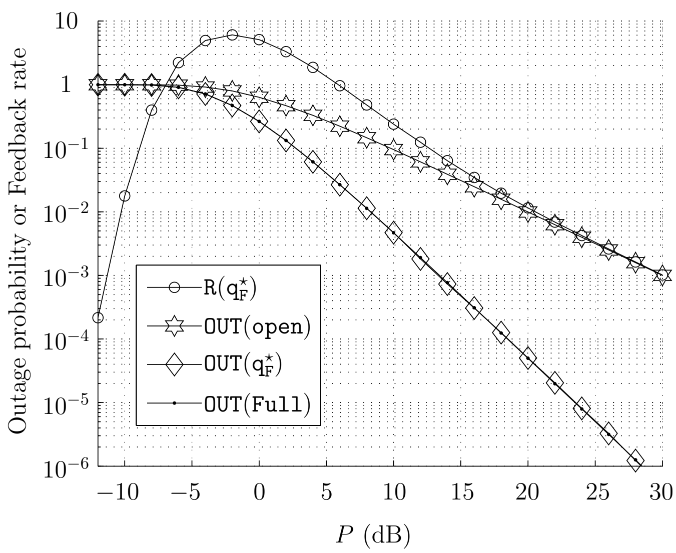


Fig. 3: The performance of the quantizer $\mathtt{q}_{\mathtt{F}}^\star$ for $t = 2$ antennas.

The performance of $\mathtt{q}_{\mathtt{F}}^\star$ for $t = 2$ antennas is as shown in Fig. 3. The horizontal axis represents P in decibels, and the vertical axis represents the outage probability or the feedback rate depending on the specific type of curve that we consider. For example, for $\mathtt{R}(\mathtt{q}_{\mathtt{F}}^\star)$, it represents the feedback rate, while for the rest of the data, it represents the outage probability.

We first discuss the outage performance of $\mathtt{q}_{\mathtt{F}}^\star$ as shown by the curve $\mathtt{OUT}(\mathtt{q}_{\mathtt{F}}^\star)$ in the figure. For comparison, we have also plotted the curve $\mathtt{OUT}(\mathtt{Full}) = \exp(-\frac{1}{P})(1+\frac{1}{P})$. We observe that the performance of both quantizers matches $\mathtt{OUT}(\mathtt{Full})$ "perfectly," with almost unnoticable differences being a result of the (necessarily) finite number of channel samples we had to consider. Indeed, not shown in the figure, we have also calculated the average distortion for each of the quantizers via the formula given in (11), and as we expect, the distortion turned out to be equal to 0 for any given value of P. We have also plotted the open-loop outage probability with beamforming, given by $\mathtt{OUT}(\mathtt{open})$.

Regarding the rate of $\mathtt{q}_{\mathtt{F}}^\star$, we can observe that it decays to 0 exponentially fast as $P \to 0$. Moreover, we observe that $\mathtt{R}(\mathtt{q}_{\mathtt{F}}^\star) \sim \frac{1}{P}$. These results verify Theorem 3.

Note that the quantizer $\mathtt{q}_{\mathtt{F}}^\star$ sacrifices rate (for the finite-P regime) in exchange for fast encoding/decoding (e.g. even the feedback binary codewords of these quantizers are assigned suboptimally). Correspondingly, we observe that the feedback rate that guarantees the full-CSIT performance can be as high as 10 bits in the low SNR regime. We expect this low-SNR "rate bump" to be much lower in the case of an optimal quantizer (the behavior of the quantization rates for very high or very-low SNR values will still remain the same in the case of an optimal quantizer). We leave such finite-SNR optimizations as future work.

VIII. VLQs for Minimum Error Probability

The previous sections have focused on the design of VLQs that minimize the outage probability of the MISO system. Outage probability is a meaningful performance measure for channel codes with infinitely large block lengths. For channel codes with finite block lengths, studying the error probability of the system is more meaningful. In this section, we discuss how to design VLQs to minimize the error probability with a generic channel code (whose codewords are to be mapped onto the transmitter antennas via quantized beamforming vectors) and maximum-likelihood decoding at the receiver.

Consider the same system model as in Section II-A with the input-output relationship as given by (2). We shall first assume that the information-bearing symbol s is a discrete random variable with a uniform distribution on the set $\{+1,-1\}$ (Note that this is a very extreme case of a channel code with a block length of 1 and a rate of 1 bit per transmission. We will later extend our results to a general class of length-ℓ rate-ρ channel codes.). For a given channel state $\mathbf{h}$, the conditional symbol error rate with a maximum-likelihood decoder is then

$$\mathtt{CSER}(\mathbf{x},\mathbf{h}) \triangleq \mathrm{Q}(\sqrt{2|\langle \mathbf{x},\mathbf{h}\rangle|^2 P}). \tag{47}$$

The (average) symbol error rate (SER) with an arbitrary mapping $\mathtt{m}: \mathbb{C}^t \to \chi$ is

$$\mathtt{SER}(\mathtt{m}) \triangleq \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|\langle \mathtt{m}(\mathbf{h}),\mathbf{h}\rangle|^2 P}\right)\right]. \tag{48}$$

In order to determine the behavior of $\mathtt{SER}(\mathtt{m})$ for large P, we let

$$\mathtt{d}(\mathtt{m}) = -\lim_{P\to\infty}\frac{\log \mathtt{SER}(\mathtt{m})}{\log P} \tag{49}$$

denote the diversity gain with $\mathtt{m}$, and

$$\mathtt{g}(\mathtt{m}) = \left[\lim_{P\to\infty}\left(\mathtt{SER}(\mathtt{m})P^{\mathtt{d}(\mathtt{m})}\right)\right]^{-1} \tag{50}$$

denote the array gain with $\mathtt{m}$, provided that both limits exist. The asymptotic $P\to\infty$ performance of $\mathtt{m}$ is then

$$\mathtt{SER}(\mathtt{m}) \sim \left[\mathtt{g}(\mathtt{m})P^{\mathtt{d}(\mathtt{m})}\right]^{-1}. \tag{51}$$

Similarly to the case of outage probability, in a full-CSIT system, the minimum SER is achieved by the mapping $\mathtt{Full}(\mathbf{h}) = \frac{\mathbf{h}^\star}{\|\mathbf{h}\|}$. This gives us

$$\mathtt{SER}(\mathtt{Full}) = \mathtt{E}[\mathrm{Q}(\sqrt{2\|\mathbf{h}\|^2 P})] \tag{52}$$

with $\mathtt{d}(\mathtt{Full}) = t$.

For the outage probability performance measure, we have designed finite-rate quantizers that can achieve the outage probability with full-CSIT. In the case of SER, a first analogous natural question to ask is whether or not the SER with full-CSIT, $\mathtt{SER}(\mathtt{Full})$, is achievable with a finite-rate quantizer. We first answer this question in the negative.

A. The Impossibility of Achieving $\mathtt{SER}(\mathtt{Full})$ with Finite-Rate Quantizers

Consider a quantizer $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_{\mathcal{I}}$ as described in Section II-C. The SER with $\mathtt{q}$ can be expressed as

$$\mathtt{SER}(\mathtt{q}) = \mathtt{SER}(\mathtt{Full}) + \mathtt{E}[\widehat{d}(\mathbf{h}, \mathtt{q}(\mathbf{h}))], \tag{53}$$

where

$$\widehat{d}(\mathbf{h}, \mathbf{x}) \triangleq \mathrm{Q}\left(\sqrt{2|\langle \mathbf{x}, \mathbf{h}\rangle|^2 P}\right) - \mathrm{Q}\left(\sqrt{2\|\mathbf{h}\|^2 P}\right) \tag{54}$$

is the distortion function associated with the SER performance measure.

We recall from (11) the distortion function $d(\mathbf{h}, \mathbf{x}) = \mathbf{1}(|\langle \mathbf{x}, \mathbf{h}\rangle|^2 < \alpha,\, \|\mathbf{h}\|^2 \geq \alpha) \in \{0, 1\}$ for the outage probability performance measure. Given $\mathbf{x}$, the distortion $d(\mathbf{h}, \mathbf{x})$ is equal to 0 on a set of channel states with positive probability. This key property of the distortion function $d(\mathbf{h}, \mathbf{x})$ has allowed us to construct finite-rate quantizers that can achieve the unquantized (full-CSIT) performance. On the other hand, for the SER performance measure, for any given $\mathbf{x}$, the distortion $\widehat{d}(\mathbf{h}, \mathbf{x}) \in [0, 1]$ is equal to 0 only on a set with probability zero. As a result, any quantization cell with positive probability necessarily incurs a positive distortion, and therefore the SER with full-CSIT is not achievable with any finite-rate quantizer. Formal calculations lead to the following theorem.

Theorem 4. *For any quantizer* $\mathtt{q}$ *with a sufficiently large feedback rate* $\mathtt{R}(\mathtt{q})$*, we have,* $\forall P \geq 0$*,*

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \mathtt{C}_3 P \exp[-\mathtt{C}_4 P R(\mathtt{q})], \tag{55}$$

where $\mathtt{C}_3,\, \mathtt{C}_4 > 0$ *are constants that are independent of* P *and* $\mathtt{R}(\mathtt{q})$.

Proof. See Appendix H. ∎

According to Theorem 4, we have no hope in achieving $\mathtt{SER}(\mathtt{Full})$ with a finite-rate quantizer. One design goal might then be to at least design quantizers that can minimize the SER for a given finite feedback rate and a given finite power constraint. However, the complicated nature of the distortion function $\widehat{d}(\mathbf{h}, \mathbf{x})$ makes the design and performance analysis of these quantizers very difficult, if not impossible. We thus focus on minimizing the SER in the $P \to \infty$ regime where the diversity and array gains of the system are the relevant performance measures. In this context, it is well-known that finite-rate FLQs cannot achieve these full-CSIT diversity and array gains [6]. In the following, we design VLQs that can achieve these gains with asymptotically zero feedback rate as $P \to \infty$. This is a significant improvement over FLQs that require infinite rate to achieve the same performance.

Before we discuss our VLQ designs in the following section, we note that one may achieve $\mathtt{d}(\mathtt{Full})$ and $\mathtt{g}(\mathtt{Full})$ while not achieving $\mathtt{SER}(\mathtt{Full})$ at any P so that our results will not contradict Theorem 4. For example, suppose that a hypothetical quantizer $\mathtt{q}'$ achieves $\mathtt{SER}(\mathtt{q}') = \mathtt{SER}(\mathtt{Full}) + 1/P^{t+1}$. Obviously we have $\mathtt{SER}(\mathtt{q}') > \mathtt{SER}(\mathtt{Full})$, $\forall P$, while $\mathtt{d}(\mathtt{q}') = \mathtt{d}(\mathtt{Full})$ and $\mathtt{g}(\mathtt{q}') = \mathtt{g}(\mathtt{Full})$.

B. The New Encoding Rule for SER-Minimizing VLQs

To achieve the goal of designing high-performance VLQs, we start with the design of the quantizer encoding regions for a given beamforming codebook $\mathcal{B}$. Our main intuition is that we do not have to pick the best beamforming vector in $\mathcal{B}$ if our goal is to achieve the diversity and array gains provided by $\mathcal{B}$. For example, we do not need to distinguish between two beamforming vectors given that both provide an SER of at most $o(1/P^t)$; preferring one vector over the other will not affect the diversity and array gains of the system as the best possible decay of the SER is $O(1/P^t)$.

With this observation, for a given beamforming codebook $\mathcal{B}$, we consider a variable-length quantizer $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ that operates as follows. Let $\beta = (t-1)\log P + g(P)$ for some $g(P) \in \omega(1) \cap O(\log P)$ (For example, one may choose $g(P) = 2\log P$ since the conditions $2\log P \in O(\log P)$ and $2\log P \in \omega(1)$ are satisfied. We may also choose $g(P) = \log\log P$.). Also, let $\mathbf{e}_i \triangleq [e_{i1} \cdots e_{it}]$ with $e_{ii} = 1$ and $e_{ij} = 0,\, j \neq i$ denote the beamforming vector that selects the ith transmitter antenna.

- If $|\langle \mathbf{e}_1, \mathbf{h}\rangle|^2 P = |h_1|^2 P \geq \beta$, then $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ feeds back the empty codeword ϵ, and we set $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h}) = \mathbf{e}_1$.
- If $|h_1|^2 P < \beta$ and $|\langle \mathbf{e}_2, \mathbf{h}\rangle|^2 P = |h_2|^2 P \geq \beta$, then $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ feeds back the binary codeword $\mathtt{0}$, and we set $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h}) = \mathbf{e}_2$.
- Otherwise, if $|h_1|^2 P < \beta$ and $|h_2|^2 P < \beta$, then $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ feeds back the concatenation of the binary codeword $\mathtt{1}$ and the binary codeword of length $\lceil \log_2 |\mathcal{B}| \rceil$ bits that represents the index, say $j \in \mathcal{I}$, of the beamforming vector in $\mathcal{B}$ that results in the maximum SNR. We set $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h}) = \mathtt{q}_{\mathcal{B}}(\mathbf{h})$.

Therefore, the variable-length quantizer $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ uses the beamforming codebook $\mathcal{B} \cup \{\mathbf{e}_1, \mathbf{e}_2\}$. We note that instead of the auxiliary vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, one can use two linearly independent vectors in $\mathcal{B}$. We incorporate the auxiliary vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ as they result in a much simpler analysis without changing the final results.

Note that the quantizer cell corresponding to the beamforming vector $\mathbf{e}_1$ does not contribute to the average feedback rate as it employs the empty codeword ϵ of length 0. On the other hand, the contribution of the vector $\mathbf{e}_2$ to the feedback rate is $\mathtt{P}(|h_1|^2 P < \beta,\, |h_2|^2 P \geq \beta) \simeq \mathtt{P}(|h_1|^2 P < \beta) \simeq \frac{\beta}{P}$. The rate contribution for each of the remaining beamforming vectors will then be in the order of $\mathtt{P}(|h_1|^2 P < \beta,\, |h_2|^2 P < \beta) \simeq \frac{\beta^2}{P^2} = o(\frac{\beta}{P})$ (provided that $\beta \in o(P)$), which results in a total average feedback rate of $\frac{\beta}{P} + o(\frac{\beta}{P})$. Hence, carefully choosing the feedback binary codewords for the first two beamforming vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ allows us to find the correct $P \to \infty$ asymptotic behavior of the feedback rate. We provide the formal derivations in the following.

C. Performance with the New Encoding Rule for Arbitrary Codebooks

We now analyze, in the following proposition, the rate and the SER performance of $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ for a general $\mathcal{B}$ and $g(P) \in \omega(1) \cap O(\log P)$.

Proposition 5. *For any finite-cardinality beamforming codebook* $\mathcal{B}$ *with* $|\mathcal{B}| \geq 2$ *and any* $g(P) \in \omega(1) \cap O(\log P)$*, we* *have*

$$\texttt{SER}(\texttt{q}^{\text{v}}_{\mathcal{B}}) \leq \texttt{SER}(\texttt{q}_{\mathcal{B}}) + \frac{2}{P^t}e^{-g(P)}, \tag{56}$$

and

$$\texttt{R}(\texttt{q}^{\text{v}}_{\mathcal{B}}) \leq \frac{(t-1)\log P}{P} + \frac{g(P)}{P} + \frac{\texttt{C}_5 \log|\mathcal{B}|\log^2 P}{P^2} \tag{57}$$

for every sufficiently large P, where $\texttt{C}_5$ is a constant that is independent of $\mathcal{B}$ and P.

Proof. See Appendix I. ■

Note that the maximum diversity gain with any quantizer is t. Hence, $\texttt{SER}(\texttt{q}_{\mathcal{B}}) \simeq g(\texttt{q}_{\mathcal{B}})P^{-d}$ for some $d \leq t$. Since $g(P) \in \omega(1)$, the second term in the upper bound in (56) decays faster than $\frac{1}{P^t}$, and thus the diversity and array gains with $\texttt{q}^{\text{v}}_{\mathcal{B}}$ is the same with those of $\texttt{q}_{\mathcal{B}}$. Moreover, according to (57), we have $\texttt{R}(\texttt{q}^{\text{v}}_{\mathcal{B}}) \to 0$ as $P \to \infty$. This is a significant improvement over a rate-$\lceil \log_2|\mathcal{B}|\rceil$ FLQ for codebook $\mathcal{B}$, especially when $|\mathcal{B}|$ is large.

We now claim that for any $f(P) \in \omega(1)$, there is a VLQ that can achieve $\texttt{d}(\texttt{Full})$ and $\texttt{g}(\texttt{Full})$ with rate $(t-1)\frac{\log P}{P} + \frac{f(P)}{P}$. We provide here an outline of the strategy to prove this result. First, we note that it is sufficient to prove the case where $f(P) \in \omega(1) \cap O(\log P)$ (as if $f(P) \in \omega(1) - O(\log P)$, we may consider quantizers with higher rates, and a higher quantization rate means a better SER performance). Then, motivated by (57), we consider P-dependent codebooks $\overline{\mathcal{B}}_P$ that satisfy

$$|\overline{\mathcal{B}}_P| = \max\left\{n : n \in \mathbb{N},\, \frac{\texttt{C}_5 \log(n)\log^2 P}{P} \leq \frac{f(P)}{2}\right\}. \tag{58}$$

and let

$$g(P) = \frac{f(P)}{2}. \tag{59}$$

Note that according to (59), we have $g(P) \in \omega(1) \cap O(\log P)$, and therefore, Proposition 5 is applicable. Moreover, our choice in (58) implies that $|\overline{\mathcal{B}}_P| \in \omega(1)$, or equivalently, we use codebooks with larger and larger cardinality as $P \to \infty$. If we can design the codebooks $\overline{\mathcal{B}}_P$, $P > 0$ well enough, an application of Proposition 5 will then reveal that the quantizer $\texttt{q}^{\text{v}}_{\overline{\mathcal{B}}_P}$ achieves $\texttt{d}(\texttt{Full})$ and $\texttt{g}(\texttt{Full})$ with rate $(t-1)\frac{\log P}{P} + \frac{f(P)}{P}$ as claimed. Obviously, for this strategy to work, we need "good" codebook designs. The existence of good codebooks have previously been established in Section IV-B via (26) and Proposition 3. For a simpler exposition, we shall use the following restatement of these results.

Proposition 6. *For every $0 < \delta < 1$, there is a codebook $\mathcal{B}_\delta$ with*

$$|\mathcal{B}_\delta| \leq \texttt{C}_6\delta^{-2t}, \tag{60}$$

and

$$\forall \mathbf{k} \in \chi,\, \exists \mathbf{x} \in \mathcal{B}_\delta,\, |\langle \mathbf{x}, \mathbf{k}\rangle|^2 \geq 1 - \delta, \tag{61}$$

where $\texttt{C}_6$ is a δ-independent constant.

Let us now calculate the SER with $\mathcal{B}_\delta$.

Proposition 7. *For every $0 < \delta \leq \frac{1}{2t}$, we have*

$$\texttt{SER}(\texttt{q}_{\mathcal{B}_\delta}) \leq \texttt{SER}(\texttt{Full})(1 + 2t\delta). \tag{62}$$

Proof. See Appendix J. ■

Hence, for every δ that satisfies $0 < \delta \leq \frac{1}{2t}$, the codebook $\mathcal{B}_\delta$ can provide the full-diversity gain. Also, with a small-enough δ, the codebook $\mathcal{B}_\delta$ can provide an array gain that is arbitrarily close to $\texttt{g}(\texttt{Full})$.

D. The Main Achievability Result

We can now proceed with the strategy outlined in Section VIII-C. The following is the main result of this section.

Theorem 5. *For every $f(P) \in \omega(1)$, there is a quantizer $\texttt{q}$ with*

$$\texttt{d}(\texttt{q}) = \texttt{d}(\texttt{Full}), \tag{63}$$

$$\texttt{g}(\texttt{q}) = \texttt{g}(\texttt{Full}), \tag{64}$$

and

$$\texttt{R}(\texttt{q}) \leq (t-1)\frac{\log P}{P} + \frac{f(P)}{P} \tag{65}$$

for all sufficiently large P.

Proof. See Appendix K. ■

Therefore, the full-CSIT diversity and array gains can be achieved with asymptotically zero feedback rate. More specifically, with the choice of e.g. $f(P) = \log\log P$, we can achieve the full-CSIT gains with rate $(t - 1 + o(1))\frac{\log P}{P}$. The question is now to determine the minimum rate that guarantees the full-CSIT gains. We discuss this problem next.

E. Necessary Conditions for Achieving $\texttt{d}(\texttt{Full})$ *and* $\texttt{g}(\texttt{Full})$

It is difficult to determine the exact asymptotic rate that guarantees the full-CSIT gains. Instead, we provide bounds. Note that by Theorem 5, a quantization rate of $(t - 1 + o(1))\frac{\log P}{P}$ is sufficient for the full-CSIT gains. In the following theorem, we prove that a quantization rate of $(t-1-o(1))\frac{\log P}{P}$ is the best possible rate that we can hope for.

Theorem 6. *For any quantizer $\texttt{q}$, if $\texttt{d}(\texttt{q}) = \texttt{d}(\texttt{Full})$, $\texttt{g}(\texttt{q}) = \texttt{g}(\texttt{Full})$, then*

$$\forall\epsilon > 0,\, \texttt{R}(\texttt{q}) \geq (t - 1 - \epsilon)\frac{\log P}{P} \tag{66}$$

for every sufficiently large P.

Proof. See Appendix L. ■

Combining the statements of Theorems 5 and 6, the necessary and sufficient feedback rate that guarantees the full-CSIT gains[1] is a member of the class of functions $(t - 1 + o(1))\frac{\log P}{P} \cup (t - 1 - o(1))\frac{\log P}{P}$. Therefore, the gap between our achievability and converse results is in the order of $o(\frac{\log P}{P})$. A tighter characterization of the necessary and sufficient feedback rate expression will remain as an open problem.

[1]For brevity of discussions, we assume that such a necessary and sufficient feedback rate exists.

F. Extensions to General Channel Codes

We have determined the necessary and sufficient feedback rates that guarantee the full-CSIT gains for the special case where binary modulation (a channel code with a block length of 1) is employed for data transmission. We now provide extensions of our results to general channel codes with possibly-P-dependent rates. We will assume maximum-likelihood decoding.

We fix some $\ell \geq 1$, and consider a $(2^{\rho\ell}, \ell)$ code[2] that can be uniquely described via its codeword alphabet $\mathcal{S} = \{\mathbf{s}_k \in \mathbb{C}^{\ell \times 1} : k = 1, \ldots, 2^{\rho\ell}\}$ with $\sum_{\mathbf{s}\in\mathcal{S}} \|\mathbf{s}\|^2 \leq \ell|\mathcal{S}|$. Although ℓ is fixed, we allow the coding rate ρ to vary with P. For a simpler discussion, we assume that $\rho \geq 1$ for every sufficiently large P. We also assume that the multiplexing gain of the system $\phi \triangleq \lim_{P\to\infty} \frac{\rho}{\log P}$ exists and satisfies $0 \leq \phi < 1$. Recall that the maximum achievable diversity gain of the MISO system for a given multiplexing gain ϕ is $t(1-\phi)$.

With this setting, the transmitter transmits a channel codeword $\mathbf{s}$ (uniformly drawn from the alphabet $\mathcal{S}$) via a beamforming vector $\mathbf{x} \in \mathbb{C}^{t\times 1}$ over ℓ time slots. In other words, at time slot $i \in \{1, \ldots, \ell\}$, the transmitter sends the signal $s_i x_j^\dagger \sqrt{P}$ over its jth antenna, where s_i is the ith component of the channel codeword $\mathbf{s}$, and x_j is the jth component of the beamforming vector $\mathbf{x}$. The channel input-output relationship with such a data transmission strategy is $\mathbf{y} = \mathbf{s}\langle \mathbf{h}, \mathbf{x}\rangle\sqrt{P} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^{\ell\times 1}$ is the received signal vector, and $\mathbf{n} \in \mathtt{CN}(\mathbf{I}_\ell)$ is the noise.

For a given channel state $\mathbf{h}$ and alphabet $\mathcal{S}$, let

$$\mathtt{CBLER}(\mathbf{x}, \mathbf{h}; \mathcal{S}) \triangleq \mathtt{P}\left(\mathbf{s} \neq \arg\min_{\mathbf{t}\in\mathcal{S}} \left\|\mathbf{y} - \mathbf{t}\langle \mathbf{h}, \mathbf{x}\rangle\sqrt{P}\right\|\right) \quad (67)$$

denote the conditional block error rate (CBLER) with a beamforming vector $\mathbf{x}$ and a maximum-likelihood decoder at the receiver. Note that the probability expression in (67) involves averaging out all possible $\mathbf{s}$ and all possible $\mathbf{n}$.

A simple exact expression for the CBLER is not available for a general $\mathcal{S}$. However, it can at least be shown that regardless of what $\mathcal{S}$ is, the CBLER decays monotonically as the SNR $|\langle \mathbf{x}, \mathbf{h}\rangle|^2 P$ increases. Hence, letting $\mathtt{BLER}(\mathtt{q}; \mathcal{S}) = \mathtt{E}_{\mathbf{h}}[\mathtt{CBLER}(\mathtt{q}(\mathbf{h}), \mathbf{h}; \mathcal{S})]$ denote the average CBLER, we have $\mathtt{BLER}(\mathtt{Full}; \mathcal{S}) \leq \mathtt{BLER}(\mathtt{q}; \mathcal{S})$ for any quantizer $\mathtt{q}$, where $\mathtt{Full}(\cdot)$ is the full-CSIT mapping. The goal is then to design a quantizer with $\mathtt{d}(\mathtt{q}) = \mathtt{d}(\mathtt{Full}; \mathcal{S})$ and $\mathtt{g}(\mathtt{q}) = \mathtt{g}(\mathtt{Full}; \mathcal{S})$, where for any given mapping $\mathtt{m} : \mathbb{C}^t \to \mathcal{X}$, we let $\mathtt{d}(\mathtt{m}; \mathcal{S})$ and $\mathtt{g}(\mathtt{m}; \mathcal{S})$ respectively denote the diversity and array gains corresponding to $\mathtt{BLER}(\mathtt{m}; \mathcal{S})$.

Regarding the CBLER with $\mathcal{S}$, we show in Appendix M that for any $\mathcal{S}$, the lower bound

$$\mathtt{CBLER}(\mathbf{x}, \mathbf{h}; \mathcal{S}) \geq 0.3\mathrm{Q}\left(\sqrt{72\ell|\langle \mathbf{x}, \mathbf{h}\rangle|^2 2^{-\rho}P}\right) \quad (68)$$

holds. In Appendix M, we also show that conversely, there is an alphabet $\mathcal{S}_0$ (roughly speaking, we set $\mathcal{S}_0$ to be the ℓth Cartesian power of a 2^ρ-QAM alphabet) whose CBLER satisfies

$$\mathtt{CBLER}(\mathbf{x}, \mathbf{h}; \mathcal{S}_0) \leq 4\ell\mathrm{Q}\left(\sqrt{8|\langle \mathbf{x}, \mathbf{h}\rangle|^2 2^{-\rho}P}\right). \quad (69)$$

[2]For a simpler presentation, we assume $2^{\rho\ell}$ is a positive integer.

From now on, we thus focus only on such non-degenerate alphabets whose CBLERs admit an upper bound of the form (69). In other words, we assume that our alphabet $\mathcal{S}$ satisfies

$$\mathtt{CBLER}(\mathbf{x}, \mathbf{h}; \mathcal{S}) \leq \mathtt{C}_7\mathrm{Q}\left(\sqrt{\mathtt{C}_8|\langle \mathbf{x}, \mathbf{h}\rangle|^2 2^{-\rho}P}\right) \quad (70)$$

for constants $\mathtt{C}_7, \mathtt{C}_8 > 0$ that are independent of ρ and P.

Using (68), (70), and the same ideas as in Section IV, we can now determine the necessary and sufficient feedback rates that guarantee the full-CSIT gains with $\mathcal{S}$. Our main result in this context is the following theorem.

Theorem 7. *There exists a quantizer* $\mathtt{q}$ *with*

$$\mathtt{d}(\mathtt{q}) = \mathtt{d}(\mathtt{Full}; \mathcal{S}) = t(1-\phi), \quad (71)$$

$$\mathtt{g}(\mathtt{q}) = \mathtt{g}(\mathtt{Full}; \mathcal{S}), \quad (72)$$

and

$$\mathtt{R}(\mathtt{q}) \leq \mathtt{C}_9 \frac{2^\rho \log P}{P} \quad (73)$$

for every sufficiently large P*, where* $\mathtt{C}_9 > 0$ *is independent of* ρ *and* P*.*

Conversely, for any quantizer $\mathtt{q}$ *that satisfies (71) and (72), we have*

$$\mathtt{R}(\mathtt{q}) \geq \mathtt{C}_{10} \frac{2^\rho \log P}{P}. \quad (74)$$

for every sufficiently large P*, where* $\mathtt{C}_{10} > 0$ *is a constant that is independent of* ρ *and* P*.*

Proof. See Appendix N. ■

Combining the main results (73) and (74) of Theorem 7, we can conclude that the necessary and sufficient feedback rate that guarantees the full-CSIT gains is

$$\Theta(1)\frac{2^\rho \log P}{P} \quad (75)$$

with the understanding that the $\Theta(1)$ term does not depend on P. In particular, setting $\rho = 1$ in (75), we obtain the slightly weakened version $\Theta(1)\frac{\log P}{P}$ of the necessary and sufficient feedback rate $(t - 1 \pm o(1))\frac{\log P}{P}$ that we have derived in Section IV for the special case of a BPSK alphabet. As a more "sophisticated" application, if we are interested in codes with a multiplexing gain of $0 \leq r < 1$, we may e.g. set $\rho = r\log_2 P$. Then, (75) tells us that the necessary and sufficient feedback rate that guarantees the full-CSIT gains with beamforming is $\Theta(1)\frac{\log P}{P^{1-r}}$. Hence, for a large class of codes and a wide range of data transmission rates, the full-CSIT gains can be achieved with asymptotically zero feedback rate by using a variable-length quantized beamforming strategy.

We note that the results of this section were derived under the assumption that one considers a fixed and finite block length ℓ. As $\ell \to \infty$, the BLER performance of the best length-ℓ rate-ρ block code should coincide with the outage probability at rate ρ. Correspondingly, as $\ell \to \infty$, we expect the performance and the structure of the BLER-optimal VLQs to "converge" to those of the outage-optimal VLQs discussed in the previous sections. Rigorously establishing such a connection will remain as an interesting open problem.

G. Numerical Results

In this section, we provide numerical evidence that supports our analytical results. In particular, we shall numerically verify our assertion that any well-designed finite-rate VLQ can achieve the full-CSIT diversity and array gains. We will also show that even at moderate-to-large transmission power levels, the SERs/BLERs with low-rate VLQs is still very close to the SERs/BLERs with full CSIT. We also compare the performance of VLQs with FLQs and show that VLQs outperform FLQs by a significant margin.

The performance of VLQs and FLQs with different feedback rates and $t = 3$ antennas is as shown in Fig. 4. The horizontal axis represents P in decibels, and the vertical axis represents the SER with a BPSK constellation. In the figure, the curve "$\mathtt{SER(open)}$" represents the SER of an open-loop system with no feedback (0 feedback bits), and "$\mathtt{SER(Full)}$" represents the SER with full CSIT (∞ feedback bits). The curves "FLQ, 1 bit" and "FLQ, 2 bits" represent the performance of the best FLQs we have found with 1 bit and 2 bits of feedback, respectively. Similarly, the curves "VLQ, b bits" for $b = 0.1, 0.5, 1, 2$ represent the best VLQs we were able to find with feedback rates $0.1, 0.5, 1$ and 2 bits per channel state, respectively. We have designed the FLQs using the generalized Lloyd algorithm [38]. In order to design the VLQs, we have used the entropy-constrained vector quantizer design algorithm in [39]. We have also utilized our structured quantizer designs whenever they provided a better performance.

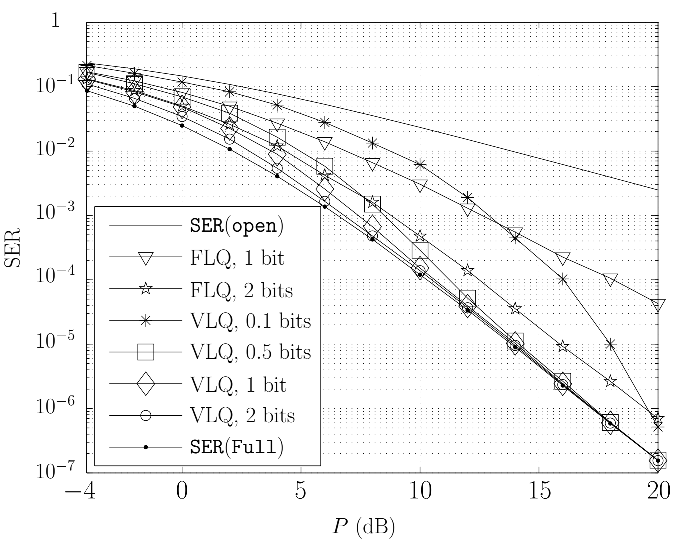


Fig. 4: SER performance of VLQs for $t = 3$ antennas and BPSK modulation.

We can observe that both the 1-bit FLQ and the 2-bit FLQ fail to achieve the full-CSIT diversity and array gains. For example, the 1-bit FLQ can only provide a diversity gain of 2. Also, while the 2-bit FLQ can achieve full diversity, it incurs an array gain loss of around 2dB compared to the full-CSIT performance. On the other hand, we can observe that the three VLQs with rates $0.5, 1$ and 2 achieve the full-CSIT diversity and array gains. The 0.1-bit VLQ will also achieve the full-CSIT gains although the convergence will be after 20dB. Still, even a 0.1-bit VLQ outperforms the 1-bit and the 2-bit FLQs when P is larger than 14dB and 20dB, respectively.

We have obtained similar results for the case of a coded modulation scheme. In Fig. 5, we show the numerical simulation results for the classical $(16, 7)$ Hamming code with a block length of 7 transmissions and a rate of $\frac{4}{7}$ bits per transmission. We have used BPSK modulation with maximum likelihood decoding over all 16 possible 7-dimensional channel codewords. The transmitter has 2 antennas. The vertical axis represents the BLER, and the horizontal axis represents P in decibels. We can observe that the open loop system can only provide a diversity gain of 1, while a full-CSIT system can achieve a diversity gain of 2. The 1-bit FLQ can achieve full diversity, it incurs an array gain loss of around 1.5dB compared to the full-CSIT performance. The two VLQs with rates 0.1 bits and 1 bit both achieve the full-CSIT gains.

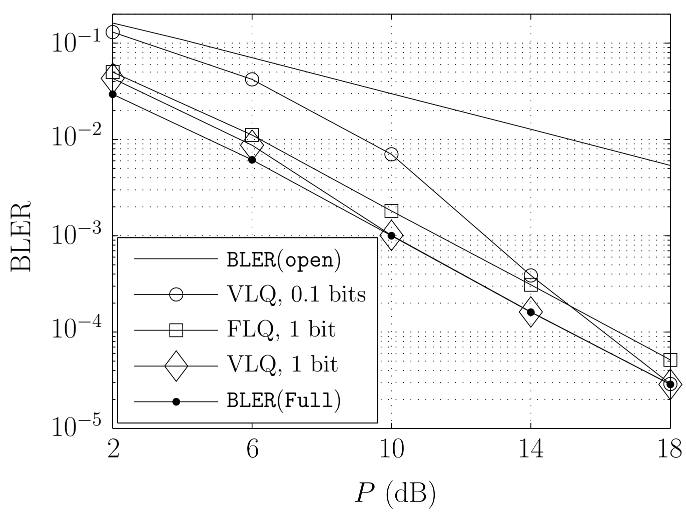


Fig. 5: BLER performance of VLQs with the $(16, 7)$ Hamming code and BPSK modulation. The transmitter has 2 antennas.

IX. Conclusions

We have considered a $t \times 1$ multiple-antenna channel with the goal of minimizing the channel outage probability or the SER by employing beamforming via quantized channel state information at the transmitter. The fact that finite-rate FLQs cannot achieve the full-CSIT outage probability performance has been previously established. We have constructed VLQs that can achieve the full-CSIT outage probability with finite rate. With P denoting the ratio of the short-term power constraint of the transmitter to the noise power at the receiver, and for a target data transmission rate of ρ bit/sec/Hz, we have shown that the necessary and sufficient VLQ rate that guarantees the full-CSIT performance is $\sim \frac{2^{\rho}-1}{P}$. We have also shown that while the SER with full-CSIT is not achievable at any finite quantization rate, the diversity and array gains of a full-CSIT system can be achieved with asymptotically zero feedback rate using variable-length quantizers. For the special case of an uncoded BPSK modulation, the necessary and sufficient feedback rate that guarantees the full-CSIT diversity and array gains is $(t-1\pm o(1))\frac{\log P}{P}$.

Our results have shown that VLQs can achieve a significantly better rate-distortion performance compared to FLQs. We believe that similar performance gains can be realized in several other quantization problems involving multiple antenna systems and similar distortion measures. For example, designing VLQs for a long-term power-constrained system is an interesting problem as future work. Analyzing the performance of VLQs under different fading models such as Nakagami fading remains as another future topic of interest. Also, the design and performance analysis of VLQs for a general MIMO system with multiple transmitter and receiver antennas remains as an open problem.

Appendix A
Proof of Proposition 2

We prove the case $\mathcal{I}=\mathbb{N}$. The case for a finite-level quantizer is very similar and thus has been skipped for brevity. First, note that for any $n\in\mathbb{N}$, we have $(\mathcal{E}_n\cap\mathcal{O}_{\mathbf{x}_n})\subset\mathcal{O}$ and $(\mathcal{E}_n\cap\mathcal{O}_{\mathbf{x}_n}^c)\subset\mathcal{O}_{\mathbf{x}_n}^c$. Therefore,

$$\begin{aligned}\mathcal{E}_n &= (\mathcal{E}_n\cap\mathcal{O}_{\mathbf{x}_n})\cup(\mathcal{E}_n\cap\mathcal{O}_{\mathbf{x}_n}^c) &(76)\\ &\subset \mathcal{O}\cup\mathcal{O}_{\mathbf{x}_n}^c. &(77)\end{aligned}$$

Hence, for an arbitrary index set $\mathcal{J}\in\{\mathbb{N},\{0\},\{0,1\},\ldots\}$, we have

$$\bigcup_{n\in\mathcal{J}}\mathcal{E}_n\subset\mathcal{O}\cup\bigcup_{n\in\mathcal{J}}\mathcal{O}_{\mathbf{x}_n}^c. \quad (78)$$

On the other hand, it can be shown by induction that the sequence $\{\mathcal{E}'_n\}_{\mathbb{N}}$ satisfies

$$\mathcal{E}'_n=\mathcal{O}_{\mathbf{x}_n}^c-\bigcup_{k=0}^{n-1}\mathcal{E}'_k,\ n\geq 1. \quad (79)$$

with $\mathcal{E}'_0=\mathcal{O}_{\mathbf{x}_0}^c\cup\mathcal{O}$. According to (21) and (22), we can then obtain

$$\bigcup_{n\in\mathcal{J}}\mathcal{E}'_n=\mathcal{O}\cup\bigcup_{n\in\mathcal{J}}\mathcal{O}_{\mathbf{x}_n}^c. \quad (80)$$

Combining (78) and (80), we arrive at

$$\bigcup_{n\in\mathcal{J}}\mathcal{E}_n\subset\bigcup_{n\in\mathcal{J}}\mathcal{E}'_n. \quad (81)$$

In particular, since $\{\mathcal{E}_n\}_{\mathbb{N}}$ are $\{\mathcal{E}'_n\}_{\mathbb{N}}$ both disjoint collections of sets, we obtain the infinite (one for each possible choice of $\mathcal{J}$) set of inequalities

$$\sum_{n\in\mathcal{J}}p_n\leq\sum_{n\in\mathcal{J}}p'_n, \quad (82)$$

where $p_n=\mathtt{P}(\mathbf{h}\in\mathcal{E}_n)$, and $p'_n=\mathtt{P}(\mathbf{h}\in\mathcal{E}'_n)$. Also, when $\mathcal{J}=\mathbb{N}$, we have $\sum_{n\in\mathbb{N}}p_n=\sum_{n\in\mathbb{N}}p'_n=1$. Without loss of generality (see Section III-C), we also have the assumption that $\ell_n\leq\ell_{n+1}$, $\forall n\in\mathbb{N}$. The rate of $\mathtt{q}$ and $\mathtt{q}'$ are given by $\mathtt{R}(\mathtt{q})=\sum_{n\in\mathbb{N}}p_n\ell_n$ and $\mathtt{R}(\mathtt{q}')=\sum_{n\in\mathbb{N}}p'_n\ell_n$, respectively.

We now rewrite the inequalities in (82) as

$$1-\sum_{n\in\mathcal{J}}p_n\geq 1-\sum_{n\in\mathcal{J}}p'_n, \quad (83)$$

or equivalently,

$$\sum_{n=N}^{\infty}p_n\geq\sum_{n=N}^{\infty}p'_n,\ N\in\mathbb{N}. \quad (84)$$

We multiply both sides of the inequality for $N=0$ by ℓ_0, and in general, for any $k\geq 1$, we multiply both sides of the inequality for $N=k$ by $(\ell_k-\ell_{k-1})$. This gives us the following infinite set of equalities and inequalities:

$$\begin{aligned}\ell_0\left(\sum_{n=0}^{\infty}p_n\right)&=\ell_0\left(\sum_{n=0}^{\infty}p'_n\right) &(85)\\ (\ell_1-\ell_0)\left(\sum_{n=1}^{\infty}p_n\right)&\geq(\ell_1-\ell_0)\left(\sum_{n=1}^{\infty}p'_n\right) &(86)\\ (\ell_2-\ell_1)\left(\sum_{n=2}^{\infty}p_n\right)&\geq(\ell_2-\ell_1)\left(\sum_{n=2}^{\infty}p'_n\right) &(87)\\ &\vdots &(88)\end{aligned}$$

By a simple telescoping series argument, the summation of the terms in the left side of these inequalities gives us $\mathtt{R}(\mathtt{q})$, while the summation of the terms in the right side is $\mathtt{R}(\mathtt{q}')$. This yields $\mathtt{R}(\mathtt{q})\geq\mathtt{R}(\mathtt{q}')$, and thus concludes the proof.

Appendix B
Proof of (29) and (30)

Let $\epsilon=\frac{1}{2^{\ell+1}}$, and $q(x)=\mathrm{sign}(x)\epsilon\lfloor|x|/\epsilon\rfloor$ (We omit the subscript of $q_\ell(\cdot)$ and write $q(\cdot)$ for brevity.). We can observe that for any $x\in[-1,1]$,

$$|q(x)|\leq|x|, \quad (89)$$

and

$$|q(x)-x|\leq\epsilon. \quad (90)$$

The last two properties (89) and (90) imply in particular that

$$|x|\leq|q(x)|+\epsilon, \quad (91)$$

by the reverse triangle inequality.

For a given $\overline{\mathbf{h}}=[\overline{h}_1\cdots\overline{h}_t]^T\in\chi$, let $\mathbf{z}=Q_\ell(\overline{\mathbf{h}})$. We have

$$\Re z_i=q(\Re\overline{h}_i),\ \Im z_i=q(\Im\overline{h}_i),\ i=1,\ldots,t. \quad (92)$$

Let us now verify (29) by showing that $0<\|\mathbf{z}\|\leq 1$. Suppose that $\|\mathbf{z}\|=0$. Then, $\mathbf{z}$ is the all-zero vector, or equivalently $q(\Re\overline{h}_i)=q(\Im\overline{h}_i)=0$, $i=1,\ldots,t$ by the definition of $\mathbf{z}$. According to (91), we then have

$$\begin{aligned}\|\overline{\mathbf{h}}\|^2&=\sum_{i=1}^{t}\left(|\Re\overline{h}_i|^2+|\Im\overline{h}_i|^2\right) &(93)\\ &\leq\sum_{i=1}^{t}\left(||q(\Re\overline{h}_i)|+\epsilon|^2+||q(\Im\overline{h}_i)|+\epsilon|^2\right) &(94)\\ &=2t\epsilon^2 &(95)\\ &=\frac{2t}{2^{2(\ell+1)}} &(96)\\ &<1, &(97)\end{aligned}$$

where the last inequality holds for $\ell \geq \lceil \log_2(2t) \rceil$ as desired. This contradicts the fact that $\|\overline{\mathbf{h}}\| = 1$. On the other hand, if $\|\mathbf{z}\| > 1$, according to (90) and (92), the inequalities

$$\begin{aligned} 1 &< \|\mathbf{z}\|^2 && (98)\\ &= \sum_{i=1}^{t} \left(|\Re z_i|^2 + |\Im z_i|^2 \right) && (99)\\ &= \sum_{i=1}^{t} \left(|q(\Re \overline{h}_i)|^2 + |q(\Im \overline{h}_i)|^2 \right) && (100)\\ &\leq \sum_{i=1}^{t} \left(|\Re \overline{h}_i|^2 + |\Im \overline{h}_i|^2 \right) && (101)\\ &= \|\overline{\mathbf{h}}\|^2 && (102)\\ &= 1 && (103) \end{aligned}$$

lead to a contradiction. We have thus established $0 < \|\mathbf{z}\| \leq 1$ and proved (29). We now verify (30) by showing that $|\langle \mathbf{z}, \overline{\mathbf{h}} \rangle|^2 > 1 - \frac{2t}{2^\ell}$. For this purpose, we first obtain a lower estimate for $\|\mathbf{z}\|^2$. By (91) and (92), we have $|\Re z_i| \geq |\Re \overline{h}_i| - \epsilon$, $|\Im z_i| \geq |\Im \overline{h}_i| - \epsilon$, $i = 1, \ldots, t$. Now, for any $i \in \{1, \ldots, t\}$, if $|\Re \overline{h}_i| \geq \epsilon$, we have

$$|\Re z_i|^2 \geq (|\Re \overline{h}_i| - \epsilon)^2 > |\Re \overline{h}_i|^2 - 2|\Re \overline{h}_i|\epsilon. \quad (104)$$

Otherwise, if $|\Re \overline{h}_i| < \epsilon$, $|\Re z_i|^2 = 0$ by the definition of $q(\cdot)$, and therefore,

$$|\Re z_i|^2 = 0 \geq |\Re \overline{h}_i|^2 - 2|\Re \overline{h}_i|\epsilon, \quad (105)$$

as the function $x^2 - 2x\epsilon$ is non-positive for all $x \in [0, 2\epsilon]$. Combining the two cases, the inequality

$$|\Re z_i|^2 \geq |\Re \overline{h}_i|^2 - 2|\Re \overline{h}_i|\epsilon \quad (106)$$

holds for any $i \in \{1, \ldots, t\}$. Noting that a similar set of inequalities holds for $|\Im z_i|^2$, $i = 1, \ldots, t$, we obtain

$$\begin{aligned} \|\mathbf{z}\|^2 &= \sum_{i=1}^{t} \left(|\Re z_i|^2 + |\Im z_i|^2 \right) && (107)\\ &> \sum_{i=1}^{t} \left(|\Re \overline{h}_i|^2 - 2|\Re \overline{h}_i|\epsilon + |\Im \overline{h}_i|^2 - 2|\Im \overline{h}_i|\epsilon \right) && (108)\\ &= 1 - 2\epsilon \sum_{i=1}^{t} \left(|\Re \overline{h}_i| + |\Im \overline{h}_i| \right). && (109) \end{aligned}$$

Subject to $\|\overline{\mathbf{h}}\|^2 = 1$, we have $\sum_{i=1}^{t} (|\Re \overline{h}_i| + |\Im \overline{h}_i|) \leq \sqrt{2t}$ with equality if and only if $|\Re \overline{h}_i| = |\Im \overline{h}_i| = \frac{1}{\sqrt{2t}}, \forall i \in \{1, \ldots, t\}$. Therefore,

$$\|\mathbf{z}\|^2 > 1 - 2\sqrt{2t}\epsilon. \quad (110)$$

Moreover, according to (90) and (92), we have

$$\|\overline{\mathbf{h}} - \mathbf{z}\|^2 \leq 2t\epsilon^2. \quad (111)$$

We can now use the decomposition

$$\begin{aligned} \|\overline{\mathbf{h}} - \mathbf{z}\|^2 &= (\overline{\mathbf{h}} - \mathbf{z})^\dagger (\overline{\mathbf{h}} - \mathbf{z}) && (112)\\ &= 1 + \|\mathbf{z}\|^2 - 2\Re\langle \overline{\mathbf{h}}, \mathbf{z} \rangle. && (113) \end{aligned}$$

Isolating $\Re\langle \overline{\mathbf{h}}, \mathbf{z} \rangle$ and then using (110) and (111), we obtain

$$\begin{aligned} \Re\langle \overline{\mathbf{h}}, \mathbf{z} \rangle &= \frac{1}{2} \left(1 + \|\mathbf{z}\|^2 - \|\overline{\mathbf{h}} - \mathbf{z}\|^2 \right) && (114)\\ &> \frac{1}{2} \left(1 + (1 - 2\sqrt{2t}\epsilon) - 2t\epsilon^2 \right) && (115)\\ &= 1 - \underbrace{\sqrt{2t}}_{\leq t}\epsilon - t\underbrace{\epsilon^2}_{<\epsilon} && (116)\\ &> 1 - 2t\epsilon. && (117) \end{aligned}$$

Therefore,

$$\begin{aligned} |\langle \overline{\mathbf{h}}, \mathbf{z} \rangle|^2 &\geq (\Re\langle \overline{\mathbf{h}}, \mathbf{z} \rangle)^2 && (118)\\ &> (1 - 2t\epsilon)^2 && (119)\\ &> 1 - 4t\epsilon. && (120) \end{aligned}$$

Substituting the value of ϵ, we obtain the desired result.

Appendix C
Proof of Theorem 1

The case $P = 0$ is trivial. Thus, suppose that $P > 0$. According to Proposition 1, for any $\ell \in \mathbb{N}$, we have

$$\mathtt{OUT}(\mathtt{q}^\star) = \mathtt{P}\left(\mathbf{h} \in \bigcap_{n \in \mathbb{N}} \mathcal{O}_{\mathbf{y}_n} \right) \leq \mathtt{P}\left(\mathbf{h} \in \bigcap_{\mathbf{y} \in \mathcal{Y}_\ell} \mathcal{O}_{\mathbf{y}} \right), \quad (121)$$

The inequality follows since $\mathcal{Y}_\ell \subset \{\mathbf{y}_n\}_{\mathbb{N}}$ by construction. Now,

$$\begin{aligned} \mathtt{P}\left(\mathbf{h} \in \bigcap_{\mathbf{y} \in \mathcal{Y}_\ell} \mathcal{O}_{\mathbf{y}} \right) &= \mathtt{P}(\forall \mathbf{y} \in \mathcal{Y}_\ell,\, |\langle \mathbf{y}, \overline{\mathbf{h}} \rangle|^2 \|\mathbf{h}\|^2 < \alpha) && (122)\\ &\leq \mathtt{P}\left(\|\mathbf{h}\|^2 < \frac{\alpha}{1 - 2t/2^\ell} \right), && (123) \end{aligned}$$

where $\overline{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|$ and the last inequality follows from Proposition 3 for sufficiently large ℓ. Since ℓ can be chosen arbitrarily large, we obtain

$$\mathtt{OUT}(\mathtt{q}^\star) \leq \mathtt{P}(\|\mathbf{h}\|^2 < \alpha) = \mathtt{OUT}(\mathtt{Full}). \quad (124)$$

Since, $\mathtt{OUT}(\mathtt{q}^\star) \geq \mathtt{OUT}(\mathtt{Full})$ is obvious, we have $\mathtt{OUT}(\mathtt{q}^\star) = \mathtt{OUT}(\mathtt{Full})$.

We need to estimate the probabilities $\mathtt{P}(\mathbf{h} \in \mathcal{F}_n)$ in order to evaluate $\mathtt{R}(\mathtt{q}^\star)$. We use the following three lemmas for this purpose.

Lemma 1. *We have* $\mathtt{P}(\mathbf{h} \in \mathcal{F}_1) \leq e^{-\alpha}\alpha$.

Proof. By the definition of $\mathcal{F}_1$, we obtain

$$\begin{aligned} \mathtt{P}(\mathbf{h} \in \mathcal{F}_1) &= \mathtt{P}\left(\mathbf{h} \in \mathcal{O}_{\mathbf{y}_1}^c \cap \mathcal{O}_{\mathbf{y}_0} \right) && (125)\\ &= \mathtt{P}\left(\mathbf{h} \in \mathcal{O}_{\mathbf{e}_2}^c \cap \mathcal{O}_{\mathbf{e}_1} \right) && (126)\\ &= \mathtt{P}(|h_2|^2 \geq \alpha,\, |h_1|^2 < \alpha) && (127)\\ &= e^{-\alpha}(1 - e^{-\alpha}) && (128)\\ &\leq e^{-\alpha}\alpha, && (129) \end{aligned}$$

which concludes the proof. ∎

Lemma 2. *For any* $n \geq 2$*, we have*

$$\mathtt{P}(\mathbf{h} \in \mathcal{F}_n) \leq \mathtt{C}_{11} e^{-\alpha}(\alpha^2 + \alpha^t), \quad (130)$$

for some constant $\mathtt{C}_{11}$ *that is independent of* α.

Proof. For any $n \geq 2$, we have

$$\mathtt{P}(\mathbf{h} \in \mathcal{F}_n) = \mathtt{P}\left(\mathbf{h} \in \mathcal{O}_{\mathbf{y}_n}^c \cap \bigcap_{k=0}^{n-1} \mathcal{O}_{\mathbf{y}_k}\right) \tag{131}$$

$$\leq \mathtt{P}\left(\mathbf{h} \in \mathcal{O}_{\mathbf{y}_n}^c \cap \mathcal{O}_{\mathbf{y}_0} \cap \mathcal{O}_{\mathbf{y}_1}\right). \tag{132}$$

Since $\{\mathbf{h} : \|\mathbf{h}\|^2 < \alpha\} \subset \mathcal{O}_{\mathbf{y}_n}$, we have $\mathcal{O}_{\mathbf{y}_n}^c \subset \{\mathbf{h} : \|\mathbf{h}\|^2 \geq \alpha\}$, and thus letting $x = |h_1|^2$, $y = |h_2|^2$, and $z = |h_3|^2 + \cdots + |h_t|^2$, we can obtain

$$\mathtt{P}(\mathbf{h} \in \mathcal{F}_n) \tag{133}$$

$$\leq \mathtt{P}\left(x + y + z \geq \alpha,\, x < \alpha,\, y < \alpha\right) \tag{134}$$

$$\leq \mathtt{P}\left(x + y + z \geq \alpha,\, x < \alpha,\, y < \alpha,\, x + y < \alpha\right) + \mathtt{P}\left(x + y + z \geq \alpha,\, x < \alpha,\, y < \alpha,\, x + y \geq \alpha\right) \tag{135}$$

$$= \mathtt{P}\left(x + y + z \geq \alpha,\, x + y < \alpha\right) + \mathtt{P}\left(x < \alpha,\, y < \alpha,\, x + y \geq \alpha\right). \tag{136}$$

The second term in (136) can be evaluated as

$$\int_0^\alpha \int_{\alpha-x}^\alpha e^{-x-y} \mathrm{d}y\mathrm{d}x = e^{-\alpha} \int_0^\alpha (1 - e^{-x})\mathrm{d}x \tag{137}$$

$$\leq \frac{\alpha^2 e^{-\alpha}}{2}. \tag{138}$$

If $t = 2$, we have $z = 0$, and the first term in (136) vanishes. If $t \geq 3$, the first term is the integral

$$\int_0^\alpha \int_0^{x-\alpha} \int_{\alpha-x-y}^\infty \frac{z^{t-3} e^{-z}}{\Gamma(t-3)} e^{-x-y} \mathrm{d}z\mathrm{d}y\mathrm{d}x \tag{139}$$

$$= \int_0^\alpha \int_0^{x-\alpha} e^{-\alpha} \sum_{i=0}^{t-3} \frac{(\alpha - x - y)^i}{\Gamma(i+1)} \mathrm{d}y\mathrm{d}x \tag{140}$$

$$\leq e^{-\alpha} \sum_{i=0}^{t-3} \frac{\alpha^i}{\Gamma(i+1)} \int_0^\alpha \int_0^{x-\alpha} \mathrm{d}y\mathrm{d}x \tag{141}$$

$$= e^{-\alpha} \sum_{i=0}^{t-3} \frac{\alpha^{i+2}}{2\Gamma(i+1)}. \tag{142}$$

Combining the cases for $t = 2$ and $t \geq 3$, we obtain the statement of the lemma. ■

Lemma 3. *There is a constant* $\ell_0 \geq 1$ *such that for all* $\ell \geq \ell_0$, *we have*

$$\sum_{n=|\mathcal{Y}_\ell|}^\infty \mathtt{P}(\mathbf{h} \in \mathcal{F}_n) \leq \frac{\mathtt{C}_{12} \alpha^t e^{-\alpha}}{2^\ell}, \tag{143}$$

where $\mathtt{C}_{12}$ *is a constant that is independent of* ℓ *and* α.

Proof. Consider an arbitrary vector $\mathbf{h}_0 \in \mathbb{C}^t$ with $\|\mathbf{h}_0\|^2 > \alpha(1 + \frac{3t}{2^\ell})$. According to Proposition 3, for sufficiently large ℓ, there is a vector $\mathbf{y} \in \mathcal{Y}_\ell$ with

$$|\langle \mathbf{y}, \mathbf{h}_0 \rangle|^2 > \|\mathbf{h}_0\|^2 \left(1 - \frac{2t}{2^\ell}\right). \tag{144}$$

Using the fact that $\|\mathbf{h}_0\|^2 > \alpha(1+\frac{3t}{2^\ell})$, we have, for sufficiently large ℓ,

$$|\langle \mathbf{y}, \mathbf{h}_0 \rangle|^2 > \alpha \left(1 + \frac{3t}{2^\ell}\right)\left(1 - \frac{2t}{2^\ell}\right) \tag{145}$$

$$= \alpha \left(1 + \frac{t}{2^\ell} - \frac{6t^2}{2^{2\ell}}\right) \tag{146}$$

$$> \alpha, \tag{147}$$

which implies $\mathbf{h}_0 \in \mathcal{O}_{\mathbf{y}}^c$. In other words, for any $\mathbf{h}$ with $\|\mathbf{h}_0\|^2 > \alpha(1 + \frac{3t}{2^\ell})$, there exists $\mathbf{y} \in \mathcal{Y}_\ell$ such that $\mathbf{h} \in \mathcal{O}_{\mathbf{y}}^c$. Therefore,

$$\left\{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}\|^2 > \alpha \left(1 + \frac{3t}{2^\ell}\right)\right\} \subset \bigcup_{\mathbf{y} \in \mathcal{Y}_\ell} \mathcal{O}_{\mathbf{y}}^c. \tag{148}$$

On the other hand,

$$\{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}\|^2 < \alpha\} \subset \mathcal{O}_{\mathbf{y}},\, \forall \mathbf{y} \in \chi, \tag{149}$$

and therefore,

$$\{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}\|^2 < \alpha\} \subset \bigcap_{i \in \mathbb{N}} \mathcal{O}_{\mathbf{y}_i}. \tag{150}$$

Now,

$$\bigcup_{i=0}^{|\mathcal{Y}_\ell|-1} \mathcal{F}_i = \bigcap_{i \in \mathbb{N}} \mathcal{O}_{\mathbf{y}_i} \cup \bigcup_{i=0}^{|\mathcal{Y}_\ell|-1} \mathcal{O}_{\mathbf{y}_i}^c \tag{151}$$

$$\supset \left\{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}\|^2 < \alpha \text{ or } \|\mathbf{h}\|^2 > \alpha \left(1 + \frac{3t}{2^\ell}\right)\right\}, \tag{152}$$

where the equality is by the definition of $\{\mathcal{F}_n\}_{\mathbb{N}}$ in (14) and (15), and the last inclusion follows from (148) and (150). This implies

$$\bigcup_{i=|\mathcal{Y}_\ell|}^\infty \mathcal{F}_i = \left(\bigcup_{i=0}^{|\mathcal{Y}_\ell|-1} \mathcal{F}_i\right)^c \tag{153}$$

$$\subset \left\{\mathbf{h} \in \mathbb{C}^t : \alpha \leq \|\mathbf{h}\|^2 \leq \alpha \left(1 + \frac{3t}{2^\ell}\right)\right\}. \tag{154}$$

Therefore,

$$\mathtt{P}\left(\mathbf{h} \in \bigcup_{i=|\mathcal{Y}_\ell|}^\infty \mathcal{F}_i\right) \leq \int_\alpha^{\alpha(1+\frac{3t}{2^\ell})} \frac{x^{t-1} e^{-x}}{\Gamma(t)} \mathrm{d}x \tag{155}$$

$$< e^{-\alpha} \int_\alpha^{\alpha(1+\frac{3t}{2^\ell})} x^{t-1} \mathrm{d}x \tag{156}$$

$$= \frac{\alpha^t e^{-\alpha}}{t} \underbrace{\left[\left(1 + \frac{3t}{2^\ell}\right)^t - 1\right]}_{\leq 1 \text{ for sufficiently large } \ell} \tag{157}$$

$$< \mathtt{C}_{12} \frac{\alpha^t e^{-\alpha}}{2^\ell}, \tag{158}$$

where the last inequality holds for sufficiently large ℓ, and $\mathtt{C}_{12}$ is a constant that is independent of ℓ. ■

We can now find an upper bound on the rate of $\mathtt{q}^\star$. We have

$$\begin{aligned}
\mathtt{R}(\mathtt{q}^\star) &= \sum_{n\in\mathbb{N}} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n)\mathtt{L}(\mathbf{b}_n^\star) && (159)\\
&= \mathtt{P}(\mathbf{h}\in\mathcal{F}_1) + \sum_{n=2}^{\infty} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n)\lfloor\log_2(n+1)\rfloor && (160)\\
&= \mathtt{P}(\mathbf{h}\in\mathcal{F}_1) + \sum_{n=2}^{|\mathcal{Y}_{\ell_0}|-1} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n)\lfloor\log_2(n+1)\rfloor + \\
&\quad \sum_{\ell=\ell_0}^{\infty}\sum_{n=|\mathcal{Y}_\ell|}^{|\mathcal{Y}_{\ell+1}|-1} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n)\lfloor\log_2(n+1)\rfloor && (161)\\
&\le \mathtt{P}(\mathbf{h}\in\mathcal{F}_1) + \lfloor\log_2|\mathcal{Y}_{\ell_0}|\rfloor \sum_{n=2}^{|\mathcal{Y}_{\ell_0}|-1} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n) \\
&\quad + \sum_{\ell=\ell_0}^{\infty} \lfloor\log_2|\mathcal{Y}_{\ell+1}|\rfloor \sum_{n=|\mathcal{Y}_\ell|}^{|\mathcal{Y}_{\ell+1}|-1} \mathtt{P}(\mathbf{h}\in\mathcal{F}_n) && (162)\\
&\le \alpha e^{-\alpha} + \mathtt{C}_{11}e^{-\alpha}(\alpha^2+\alpha^t)\lfloor\log_2|\mathcal{Y}_{\ell_0}|\rfloor(|\mathcal{Y}_{\ell_0}|-2) + \\
&\quad \mathtt{C}_{12}\alpha^t e^{-\alpha}\sum_{\ell=\ell_0}^{\infty}\lfloor\log_2|\mathcal{Y}_{\ell+1}|\rfloor/2^\ell, && (163)
\end{aligned}$$

where for the first inequality, we have used the monotonicity of $\log_2(n+1)$, and for the second inequality we have applied Lemmas 1, 2 and 3. The upper bound $|\mathcal{Y}_\ell| \le 2^{2t(\ell+3)}$ in Section IV-B implies that the sum in the last inequality is finite. This concludes the proof.

APPENDIX D
PROOF OF THEOREM 2

We first prove the lower bound on $\mathtt{OUT}(\mathtt{q})$. It trivially holds if $\mathtt{R}(\mathtt{q}) \ge 1$. Therefore, suppose that $\mathtt{R}(\mathtt{q}) < 1$. Let $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_{\mathcal{I}}$. It is straightforward to show by contradiction that there is an index $i\in\mathcal{I}$ such that $\mathtt{P}(\mathbf{h}\in\mathcal{E}_i) \ge 1-\mathtt{R}(\mathtt{q})$. Without loss of generality, suppose that $\mathtt{P}(\mathbf{h}\in\mathcal{E}_1)\ge 1-\mathtt{R}(\mathtt{q})$ and $1\in\mathcal{I}$. Then, with $f(\mathbf{h})$ representing the probability density function of $\mathbf{h}$, we have

$$\begin{aligned}
\mathtt{OUT}(\mathtt{q}) &= \sum_{i\in\mathcal{I}}\int_{\mathcal{E}_i}\mathbf{1}(|\langle\mathbf{x}_i,\mathbf{h}\rangle|^2<\alpha)f(\mathbf{h})\mathrm{d}\mathbf{h} && (164)\\
&\ge \int_{\mathcal{E}_1}\mathbf{1}(|\langle\mathbf{x}_1,\mathbf{h}\rangle|^2<\alpha)f(\mathbf{h})\mathrm{d}\mathbf{h} && (165)\\
&= \underbrace{\int_{\mathbb{C}^T}\mathbf{1}(|\langle\mathbf{x}_1,\mathbf{h}\rangle|^2<\alpha)f(\mathbf{h})\mathrm{d}\mathbf{h}}_{=\mathtt{OUT}(\mathtt{open})} - \\
&\quad \int_{\mathcal{E}_1^c}\mathbf{1}(|\langle\mathbf{x}_1,\mathbf{h}\rangle|^2<\alpha)f(\mathbf{h})\mathrm{d}\mathbf{h} && (166)\\
&\ge \mathtt{OUT}(\mathtt{open}) - \int_{\mathcal{E}_1^c} f(\mathbf{h})\mathrm{d}\mathbf{h} && (167)\\
&\ge \mathtt{OUT}(\mathtt{open}) - \mathtt{R}(\mathtt{q}), && (168)
\end{aligned}$$

and this concludes the proof.

APPENDIX E
PROOF OF COROLLARY 1

For notational convenience, let $r^\star = \mathtt{r}^\star(\mathtt{OUT}(\mathtt{Full}))$. According to Theorem 1, for the quantizer $\mathtt{q}^\star$ that achieves the full-CSIT performance, we have $\mathtt{R}(\mathtt{q}^\star) \le \frac{2^\rho-1}{P} + O(\frac{1}{P^2})$, which means $r^\star \le \frac{2^\rho-1}{P} + O(\frac{1}{P^2})$. The last inequality implies $\limsup_{P\to\infty}(r^\star/\frac{2^\rho-1}{P}) \le 1$. To complete the proof, it is now sufficient to show that $\liminf_{P\to\infty}(r^\star/\frac{2^\rho-1}{P}) \ge 1$. Assume the contrary. Then, $\exists r < 2^\rho-1, \forall P_0\in\mathbb{R}, \exists P > P_0$ such that $\mathtt{r}^\star(\mathtt{OUT}(\mathtt{Full})) \le \frac{r}{P}$, or equivalently,

$$\begin{aligned}
&\exists r < 2^\rho - 1,\ \forall P_0\in\mathbb{R},\ \exists P > P_0,\ \exists \mathtt{q} \text{ such that}\\
&\qquad \mathtt{OUT}(\mathtt{q}) = \mathtt{OUT}(\mathtt{Full}),\ \mathtt{R}(\mathtt{q}) \le \frac{r}{P}. && (169)
\end{aligned}$$

Now, for any quantizer $\mathtt{q}$ in (169), we have

$$\begin{aligned}
\mathtt{OUT}(\mathtt{Full}) &= \mathtt{OUT}(\mathtt{q}) && (170)\\
&\ge \mathtt{OUT}(\mathtt{open}) - \mathtt{R}(\mathtt{q}) && (171)\\
&\ge \frac{2^\rho-1-r}{P} - \frac{1}{2}\left(\frac{2^\rho-1}{P}\right)^2 && (172)
\end{aligned}$$

where the first inequality follows from Theorem 2, and the second inequality follows from the lower bound $\mathtt{OUT}(\mathtt{open}) = 1 - e^{-\frac{2^\rho-1}{P}} \ge \frac{2^\rho-1}{P} - \frac{1}{2}(\frac{2^\rho-1}{P})^2$, and the fact that $\mathtt{R}(\mathtt{q}) \le \frac{r}{P}$. Since $1-r>0$, the lower bound in (172) is $\Theta(\frac{1}{P})$ and is strictly greater than $\mathtt{OUT}(\mathtt{Full}) \in \Theta(\frac{1}{P^t})$ when P is sufficiently large. This leads to the contradiction $\mathtt{OUT}(\mathtt{Full}) > \mathtt{OUT}(\mathtt{Full})$.

APPENDIX F
PROOF OF PROPOSITION 4

First, note that our choice of $\ell^\star$ satisfies

$$\|\mathbf{h}\|^2 \ge \alpha\left(1+\frac{4t}{2^{\ell^\star}}\right), \qquad (173)$$

and

$$\ell^\star \ge \lceil\log_2(4t)\rceil. \qquad (174)$$

Obviously, the latter implies $\ell^\star \ge \lceil\log_2(2t)\rceil$. Then, according to (30), we have

$$|\langle Q_{\ell^\star}(\overline{\mathbf{h}}), \overline{\mathbf{h}}\rangle|^2 > 1 - \frac{2t}{2^{\ell^\star}}. \qquad (175)$$

Combining with (173), we obtain

$$\begin{aligned}
|\langle Q_{\ell^\star}(\overline{\mathbf{h}}), \mathbf{h}\rangle|^2 &\ge \alpha\left(1+\frac{4t}{2^{\ell^\star}}\right)\left(1-\frac{2t}{2^{\ell^\star}}\right) && (176)\\
&= \alpha\left(1+\frac{2t}{2^{\ell^\star}} - \frac{2t}{2^{\ell^\star}} \times \underbrace{(4t/2^{\ell^\star})}_{\le 1 \text{ by } (174)}\right) && (177)\\
&\ge \alpha, && (178)
\end{aligned}$$

This implies $|\langle \overline{Q}_{\ell^\star}(\overline{\mathbf{h}}), \mathbf{h}\rangle|^2 \ge \alpha$ and thus concludes the proof.

APPENDIX G
PROOF OF THEOREM 3

We first show that $\mathtt{OUT}(\mathtt{q}_{\mathtt{F}}^\star) = \mathtt{OUT}(\mathtt{Full})$. According to our discussion Section V-A1, an outage event can only happen if the feedback codeword is ϵ. Note that the encoder feeds back ϵ if $|h_1|^2 \geq \alpha$ or $\|\mathbf{h}\|^2 \leq \alpha$, in which case the transmitter uses the beamforming vector $\mathbf{e}_1$. The case $|h_1|^2 \geq \alpha$ then does not result in any outage. Therefore, $\mathtt{OUT}(\mathtt{q}_{\mathtt{F}}^\star) \leq \mathtt{P}(\|\mathbf{h}\|^2 \leq \alpha) = \mathtt{OUT}(\mathtt{Full})$, which implies $\mathtt{OUT}(\mathtt{q}_{\mathtt{F}}^\star) = \mathtt{OUT}(\mathtt{Full})$.

For the rate of $\mathtt{q}_{\mathtt{F}}^\star$, note that the quantizer encoder feeds back $\mathtt{0}$ with probability no more than

$$\mathtt{P}(|h_1|^2 < \alpha,\, |h_2|^2 \geq \alpha) = e^{-\alpha}(1-e^{-\alpha}) \leq \alpha e^{-\alpha} \tag{179}$$

We now consider the feedback binary codewords that will be fed back at Line 7 of the encoding algorithm. Each of them has length $2t(\ell^\star+3)$ bits, where $\ell^\star \geq \lceil \log_2(4t) \rceil$ according to (42). We have $\ell^\star = \lceil \log_2(4t) \rceil$ with probability at most

$$\mathtt{P}(|h_1|^2 < \alpha,\, |h_2|^2 < \alpha,\, \|\mathbf{h}\|^2 \geq \alpha) \leq \mathtt{C}_{11} e^{-\alpha}(\alpha^2 + \alpha^t), \tag{180}$$

due to Lines 1 and 3 of Algorithm 1. The last inequality follows from Lemma (2).

On the other hand, for any $n \geq 1$, if $\ell^\star = \lceil \log_2(4t) \rceil + n$, then $\|\mathbf{h}\|^2 > \alpha$ and

$$\lceil \log_2(4t) \rceil + n \leq \left\lceil -\log_2\left(\frac{1}{4t}\left(\frac{\|\mathbf{h}\|^2}{\alpha} - 1\right)\right)\right\rceil \tag{181}$$

$$\Longrightarrow \lceil \log_2(4t) \rceil + n \leq -\log_2\left(\frac{1}{4t}\left(\frac{\|\mathbf{h}\|^2}{\alpha} - 1\right)\right) + 1 \tag{182}$$

$$\Longrightarrow \|\mathbf{h}\|^2 \leq \alpha\left(1 + \frac{\mathtt{C}_{13}}{2^n}\right), \tag{183}$$

where $\mathtt{C}_{13}$ is a constant that is independent of α and n. This implies

$$\begin{aligned} &\mathtt{P}(\ell^\star = \lceil \log_2(4t) \rceil + n) \\ &\quad \leq \mathtt{P}\left[\alpha < \|\mathbf{h}\|^2 \leq \alpha\left(1 + \frac{\mathtt{C}_{13}}{2^n}\right)\right] \quad (184) \\ &\quad \leq \frac{\mathtt{C}_{14}\alpha^t e^{-\alpha}}{2^{nt}},\, n \geq 1, \quad (185) \end{aligned}$$

where $\mathtt{C}_{14}$ is another constant that is independent of α and n. Combining (179), (180) and (185) we obtain

$$\begin{aligned} \mathtt{R}(\mathtt{q}_{\mathtt{F}}^\star) &\leq \alpha e^{-\alpha} + 2t(\lceil \log_2(4t) \rceil + 3)\mathtt{C}_{11} e^{-\alpha}(\alpha^2 + \alpha^t) + \\ &\quad \sum_{n=1}^{\infty} 2t(\log_2(4t) + n + 3)\frac{\mathtt{C}_{14}\alpha^t e^{-\alpha}}{2^{nt}}, \end{aligned} \tag{186}$$

which gives us the same upper bound in the statement of the theorem.

APPENDIX H
PROOF OF THEOREM 4

We need the following lemma.

Lemma 4. *For any quantizer $\mathtt{q} = \{p_n, \mathcal{E}_n, \mathtt{b}_n\}_{\mathcal{I}}$ with a sufficiently large rate $\mathtt{R}(\mathtt{q})$, there is an index $n \in \mathcal{I}$ such that $\mathtt{P}(X \in \mathcal{E}_n) \geq 2^{-5\mathtt{R}(\mathtt{q})}$.*

Proof. Consider the sequence of binary codewords $\bar{\mathtt{b}}_0 = \epsilon$, $\bar{\mathtt{b}}_1 = \mathtt{0}$, $\bar{\mathtt{b}}_1 = \mathtt{1}$, $\bar{\mathtt{b}}_2 = \mathtt{00}, \ldots$ that form the non-prefix-free code alphabet $\{0,1\}^\star \triangleq \{\epsilon, \mathtt{0}, \mathtt{1}, \mathtt{00}, \mathtt{01}, \ldots\}$, where ϵ is the empty codeword of length 0. We have $\mathtt{l}(\bar{\mathtt{b}}_n) = \lfloor \log_2(n+1) \rfloor$, and

$$\begin{aligned} \mathtt{R}(\mathtt{q}) &\geq \inf_{\substack{\{\ell_n\}_{\mathcal{I}} \subset \{0,1\}^\star \\ \sum_{n\in\mathcal{I}} 2^{-\ell_n} \leq 1}} \sum_{n\in\mathcal{I}} \mathtt{P}(X \in \mathcal{E}_n)\ell_n \quad (187) \\ &\geq \inf_{\{\ell_n\}_{\mathcal{I}} \subset \{0,1\}^\star} \sum_{n\in\mathcal{I}} \mathtt{P}(X \in \mathcal{E}_n)\ell_n \quad (188) \\ &= \sum_{n\in\mathcal{I}} \mathtt{P}(X \in \mathcal{E}_n)\lfloor \log_2(n+1) \rfloor, \quad (189) \end{aligned}$$

where for (189), we have assumed (without loss of generality) that $\mathtt{P}(X \in \mathcal{E}_{n-1}) \leq \mathtt{P}(X \in \mathcal{E}_n)$, $\forall n \in \mathcal{I} - \{0\}$. We now prove the lemma by contradiction. Suppose that $\mathtt{P}(X \in \mathcal{E}_n) \leq \delta$, $\forall n \in \mathcal{I}$, where $\delta \triangleq 2^{-5\mathtt{R}(\mathtt{q})}$. Optimizing (189) over all possible probabilities $\mathtt{P}(X \in \mathcal{E}_n)$, $n \in \mathcal{I}$, we obtain

$$\mathtt{R}(\mathtt{q}) \geq \inf_{\substack{0 \leq p_n \leq \delta,\, \forall n\in\mathcal{I} \\ \sum_{i\geq 1} p_i = 1-\epsilon}} \sum_{n\in\mathcal{I}} p_n \lfloor \log_2(1+n) \rfloor. \tag{190}$$

To evaluate this lower bound, we first observe that $\lfloor \log_2(1+n) \rfloor$, $n = 0, 1, \ldots,$ is a monotonically non-decreasing sequence of integers that are independent of p_n. We thus start from $n = 0$, and assign p_0 the highest possible feasible probability, i.e. $p_0 = \delta$, and continue sequentially in the same manner for indices $1, \ldots, N$, until $\sum_{n=0}^N p_n = 1$. Therefore, the infimum can be achieved, and the minimizing solution is

$$p_n^\star \triangleq \begin{cases} \delta, & 0 \leq n < \lfloor \frac{1-\epsilon}{\delta} \rfloor \\ 1 - \epsilon - \delta\lfloor \frac{1-\epsilon}{\delta} \rfloor, & n = \lfloor \frac{1-\epsilon}{\delta} \rfloor \\ 0, & n > \lfloor \frac{1-\epsilon}{\delta} \rfloor \end{cases}. \tag{191}$$

This gives us

$$\begin{aligned} \mathtt{R}(\mathtt{q}) &\geq \delta \sum_{n=0}^{\lfloor \frac{1-\epsilon}{\delta} \rfloor - 1} \lfloor \log_2(1+n) \rfloor \quad (192) \\ &\geq \delta \sum_{n=0}^{\lfloor \frac{1-\epsilon}{\delta} \rfloor - 1} (\log_2(1+n) - 1) \quad (193) \\ &= \delta \log_2\left(\lfloor \tfrac{1-\epsilon}{\delta} \rfloor !\right) - \delta \lfloor \tfrac{1-\epsilon}{\delta} \rfloor \quad (194) \\ &\geq \delta \log_2\left(\lfloor \tfrac{1-\epsilon}{\delta} \rfloor !\right) - 1. \quad (195) \end{aligned}$$

Noting that at least $\left\lfloor \frac{1}{2}\left\lfloor \frac{1}{\delta} \right\rfloor \right\rfloor = \left\lfloor \frac{1}{2\delta} \right\rfloor$ terms of the product $1 \cdot 2 \cdots \left\lfloor \frac{1}{\delta} \right\rfloor = \left\lfloor \frac{1}{\delta} \right\rfloor !$ are greater than or equal to $\frac{1}{2}\left\lfloor \frac{1}{\delta} \right\rfloor$, we obtain

$$\begin{aligned} \mathtt{R}(\mathtt{q}) &\geq \delta \left\lfloor \frac{1-\epsilon}{2\delta} \right\rfloor \log_2\left(\frac{1}{2}\left\lfloor \frac{1-\epsilon}{\delta} \right\rfloor\right) - 1 \quad (196) \\ &\geq \frac{1}{4}\log_2 \frac{1}{64\delta}, \quad (197) \end{aligned}$$

where the last inequality follows since $\left\lfloor \frac{1}{2\delta} \right\rfloor \leq \frac{1}{4\delta}$ and $\frac{1}{2}\left\lfloor \frac{1}{\delta} \right\rfloor \leq \frac{1}{4\delta}$ when $\frac{\delta}{1-\epsilon} \leq \frac{1}{4}$. The last inequality leads to a contradiction once we substitute $\delta = 2^{-5\mathtt{R}(\mathtt{q})}$, and this concludes the proof. ■

We now prove the theorem. Given $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_{\mathcal{I}}$, we have

$$\mathtt{SER}(\mathtt{q}) = \mathtt{SER}(\mathtt{Full}) + \sum_{i\in\mathcal{I}} \int_{\mathcal{E}_i} \widetilde{d}(\mathbf{x}_i, \mathbf{h}) f(\mathbf{h}) \mathrm{d}\mathbf{h}. \tag{198}$$

Let $r = 2^{-5\mathtt{R}(\mathtt{q})}$. According to Lemma 4, there is an index $i \in \mathcal{I}$ such that $\mathtt{P}(\mathbf{h} \in \mathcal{E}_i) \geq r$. Without loss of generality, suppose that $0 \in \mathcal{I}$ and $\mathtt{P}(\mathbf{h} \in \mathcal{E}_0) \geq r$. We have

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \int_{\mathcal{E}_0} \widetilde{d}(\mathbf{x}_0, \mathbf{h}) f(\mathbf{h}) \mathrm{d}\mathbf{h} \tag{199}$$

$$\geq \mathtt{SER}(\mathtt{Full}) + \inf_{\mathcal{E} \subset \mathfrak{E}} \inf_{\mathbf{x} \in \chi} g(\mathcal{E}, \mathbf{x}) \tag{200}$$

where $g(\mathcal{E}, \mathbf{x}) = \int_{\mathcal{E}} \widetilde{d}(\mathbf{x}, \mathbf{h}) f(\mathbf{h}) \mathrm{d}\mathbf{h}$ and $\mathfrak{E} = \{\mathcal{E} \subset \mathbb{C}^t : \mathtt{P}(\mathbf{h} \in \mathcal{E}) \geq r$. Note that for any $t \times t$ unitary matrix $\mathbf{U}$, we have $g(\mathcal{E}, \mathbf{x}) = g(\mathbf{U}\mathcal{E}, \mathbf{U}\mathbf{x})$, $\forall \mathcal{E} \in \mathfrak{E}$, $\forall \mathbf{x} \in \chi$, where $\mathbf{U}\mathcal{E} = \{\mathbf{U}\mathbf{h} : \mathbf{h} \in \mathcal{E}\}$ denotes the translate of the set $\mathcal{E}$ by $\mathbf{U}$. With this property in mind, we consider a fixed vector $\mathbf{y} \in \chi$. For a given $\mathbf{x} \in \chi$, let the unitary matrix $\mathbf{U}_{\mathbf{x}}$ satisfy $\mathbf{x} = \mathbf{U}_{\mathbf{x}}\mathbf{y}$. We have $g(\mathcal{E}, \mathbf{x}) = g(\mathbf{U}_{\mathbf{x}}^{\dagger}\mathcal{E}, \mathbf{U}_{\mathbf{x}}^{\dagger}\mathbf{x}) = g(\mathbf{U}_{\mathbf{x}}^{\dagger}\mathcal{E}, \mathbf{y})$. Since $\mathbf{U}_{\mathbf{x}}^{\dagger}\mathcal{E} \in \mathfrak{E}$, we have $g(\mathcal{E}, \mathbf{x}) \geq \inf_{\mathcal{E}' \in \mathfrak{E}} g(\mathcal{E}', \mathbf{y})$. Since this inequality holds for arbitrary $\mathbf{x}$ and $\mathcal{E}$, we obtain $\inf_{\mathcal{E} \in \mathfrak{E}} \inf_{\mathbf{x} \in \chi} g(\mathcal{E}, \mathbf{x}) \geq \inf_{\mathcal{E} \in \mathfrak{E}} g(\mathcal{E}, \mathbf{y})$. In particular, choosing $\mathbf{y} = [1\, 0 \cdots 0]^T$ then gives us

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \inf_{\mathcal{E} \subset \mathfrak{E}} \int_{\mathcal{E}} \left[\mathrm{Q}\left(\sqrt{2|h_1|^2 P}\right) - \mathrm{Q}\left(\sqrt{2\|\mathbf{h}\|^2 P}\right) \right] f(\mathbf{h}) \mathrm{d}\mathbf{h} \tag{201}$$

We now find a lower bound for the integrand in (201). Note that

$$\frac{\partial}{\partial x} \mathrm{Q}(\sqrt{2x}) = -\frac{e^{-x}}{2\sqrt{\pi}\sqrt{x}}, \tag{202}$$

and

$$\frac{\partial^2}{\partial x^2} \mathrm{Q}(\sqrt{2x}) = \frac{e^{-x}(2x+1)}{4\sqrt{\pi} x^{3/2}}. \tag{203}$$

Since $\frac{\partial^2}{\partial x^2} \mathrm{Q}(\sqrt{2x}) \geq 0, \forall x \geq 0$, $\mathrm{Q}(\sqrt{2x})$ is convex, and therefore, for any $x, y \geq 0$,

$$\mathrm{Q}(\sqrt{2x}) - \mathrm{Q}(\sqrt{2(x+y)}) \geq -y \frac{\partial \mathrm{Q}(\sqrt{2u})}{\partial u}\bigg|_{u=x+y} \tag{204}$$

$$= \frac{y e^{-(x+y)}}{2\sqrt{\pi}\sqrt{x+y}} \tag{205}$$

$$\geq \frac{1}{4} y e^{-2(x+y)}, \tag{206}$$

where the last inequality follows since $e^{-x}/\sqrt{x} \geq e^{-2x}$ and $\frac{1}{2\sqrt{\pi}} \geq \frac{1}{4}$.

Applying (206) to (201), we have

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \frac{P}{4} \inf_{\mathcal{E} \subset \mathfrak{E}} \int_{\mathcal{E}} \sum_{k=2}^{t} |h_k|^2 e^{-2\|\mathbf{h}\|^2 P} f(\mathbf{h}) \mathrm{d}\mathbf{h} \tag{207}$$

To further simplify the integrand and the calculation of the infimum, we define a real number y_0 that satisfies $\int_0^{y_0} \frac{y^{t-2} e^{-y}}{\Gamma(t-1)} \mathrm{d}y = \frac{r}{2}$. We have

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \frac{y_0 P}{4} \inf_{\mathcal{E} \subset \mathfrak{E}} \int_{\mathcal{E}} \mathbf{1}\left(\sum_{k=2}^{t} |h_k|^2 \geq y_0 \right) e^{-2\|\mathbf{h}\|^2 P} f(\mathbf{h}) \mathrm{d}\mathbf{h} \tag{208}$$

The expression in the lower bound is an optimization problem of the form "Minimize $\int_{\mathcal{E}} g(x) f(x) \mathrm{d}x$, subject to $\int_{\mathcal{E}} f(x) \mathrm{d}x \geq r$," where $f(x)$ is a continuous probability density function, $g(x)$ is an arbitrary measurable function. According to [40], a solution set $\mathcal{E}^\star$ exists, and one forms $\mathcal{E}^\star$ by starting with the points where the integrand $g(x)$ takes its minimal values and then progressively adds more points until the measure of $\mathcal{E}^\star$ is equal to r. In order to solve (208), we thus need to evaluate the level sets of the function $g(\mathbf{h}) = \mathbf{1}\left(\sum_{k=2}^{t} |h_k|^2 \geq y_0\right) e^{-2\|\mathbf{h}\|^2 P}$. Note that the level sets of $g(\mathbf{h})$ depend only on $X \triangleq |h_1|^2$ and $Y = \sum_{k=2}^{t} |h_k|^2$. Moreover, the PDF $f(\mathbf{h})$ can also be expressed in terms of X and Y only. Hence, we may equivalently consider

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \frac{y_0 P}{4} \inf_{\mathcal{F}} \int_{\mathcal{F}} g(x,y) f_X(x) f_Y(y) \mathrm{d}x \mathrm{d}y \tag{209}$$

such that

$$\int_{\mathcal{F}} f_X(x) f_Y(y) \mathrm{d}x \mathrm{d}y \geq r \tag{210}$$

with

$$g(x,y) = \mathbf{1}(y \geq y_0) e^{-2(x+y)P}, \tag{211}$$

$$f_X(x) = e^{-x}, \tag{212}$$

and

$$f_Y(y) = \frac{y^{t-2} e^{-y}}{\Gamma(t-1)}. \tag{213}$$

Now, let $\mathcal{F}^\star$ denote a minimizer of (209). We observe that the function $g(x,y)$ takes its minimum value 0 whenever $y < y_0$. The set of channel states with $y < y_0$ has probability $\frac{r}{2}$ by our choice of the constant y_0. Since $\frac{r}{2} < r$, we may set $\{[x,y] \in \mathbb{R}^2_{\geq 0} : y < y_0\} \subset \mathcal{F}^\star$. This leaves us with a set of probability $r - \frac{r}{2} = \frac{r}{2}$ that we need to assign to $\mathcal{F}^\star$. When $y \geq y_0$, we have $g(x,y) = e^{-2(x+y)P}$, and thus (i) $g(x,y)$ is a constant whenever $x+y$ is, and (ii) $g(x,y)$ decreases whenever $x+y$ increases. Hence, the solution set $\mathcal{F}^\star$ can be expressed as

$$\mathcal{F}^\star = \{[x,y] \in \mathbb{R}^2_{\geq 0} : y < y_0 \text{ or } x+y \geq b\}, \tag{214}$$

where b should be chosen to satisfy $\mathtt{P}([x,y] \in \mathcal{F}^\star) = r$.

To proceed further, we now find estimates on b. First, note that $b > y_0$, as otherwise if $b \leq y_0$, we have $\mathcal{F}^\star = \mathbb{R}^2_{\geq 0}$ and this leads to the contradiction $1 = \mathtt{P}(\mathbf{h} \in \mathcal{F}^\star) = r < 1$. Since $b > y_0$, $\{[x,y] \in \mathbb{R}^2_{\geq 0} : y \geq b\} \subset \mathcal{F}^\star$. Also, when $y \geq b > y_0$, we have $g(x,y) = e^{-2(x+y)P}$. Therefore,

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \frac{y_0 P}{4} \int_0^\infty \int_b^\infty e^{-2(x+y)P} e^{-x} \frac{y^{t-2} e^{-y}}{\Gamma(t-1)} \mathrm{d}x \mathrm{d}y \tag{215}$$

$$= \mathtt{SER}(\mathtt{Full}) + \frac{y_0 P}{\Gamma(t-1)(1+2P)} \int_b^\infty y^{t-2} e^{-y(1+2P)} \mathrm{d}y \tag{216}$$

Now, since $\mathtt{P}(X+Y \geq b) \leq \mathtt{P}(Y < y_0 \text{ or } X+Y \geq b) = r = 2^{-5\mathtt{R}(\mathtt{q})}$, we have (say) $b \geq 1$ for a sufficiently large feedback rate $\mathtt{R}(\mathtt{q})$. This implies $y^{t-2} \geq b^{t-2} \geq 1$ for the integrand in (216), and therefore,

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{SER}(\mathtt{Full}) + \frac{y_0 P \exp[-b(1+2P)]}{\Gamma(t-1)(1+2P)^2}. \quad (217)$$

To conclude the proof, we need to find a lower bound for y_0 and an upper bound for b. We obtain the lower bound on y_0 as

$$\frac{r}{2} = \int_0^{y_0} \underbrace{y^{t-2}}_{\leq y_0^{t-2}} \underbrace{e^{-y}}_{\leq 1} \underbrace{(\Gamma(t-1))^{-1}}_{\leq 1} \mathrm{d}y \leq y_0^{t-1} \implies$$
$$y_0 \geq 2^{-\frac{1}{t-1}} r^{\frac{1}{t-1}} \geq \frac{1}{2} r^{\frac{1}{t-1}}. \quad (218)$$

In order to obtain an upper bound for b, we first note that

$$\begin{aligned} r &= \mathtt{P}(Y < y_0 \text{ or } X + Y \geq b) && (219)\\ &\leq \mathtt{P}(Y < y_0) + \mathtt{P}(X + Y \geq b) && (220)\\ &= \frac{r}{2} + \mathtt{P}(X + Y \geq b), && (221)\end{aligned}$$

and therefore

$$\begin{aligned} \frac{r}{2} &\leq \mathtt{P}(X + Y \geq b) && (222)\\ &= \int_b^{\infty} \frac{x^{t-1} \overbrace{e^{-x}}^{\leq e^{-\frac{x}{2}} e^{-\frac{b}{2}}}}{\Gamma(t)} \mathrm{d}x && (223)\\ &\leq e^{-\frac{b}{2}} \int_b^{\infty} \frac{x^{t-1} e^{-\frac{x}{2}}}{\Gamma(t)} \mathrm{d}x && (224)\\ &\leq e^{-\frac{b}{2}} \int_0^{\infty} \frac{x^{t-1} e^{-\frac{x}{2}}}{\Gamma(t)} \mathrm{d}x && (225)\\ &= 2^t e^{-\frac{b}{2}}, && (226)\end{aligned}$$

which implies $b \leq -2 \log \frac{r}{2^{t+1}}$. Substituting the bounds for y_0 and b, and $r = 2^{-5\mathtt{R}(\mathtt{q})}$ to (217), we obtain the same lower bound as in the statement of the theorem after some straightforward manipulations. This concludes the proof.

Appendix I
Proof of Proposition 5

We first prove the upper bound on $\mathtt{SER}(\mathtt{q}_{\mathcal{B}}^{\mathrm{v}})$. Let

$$\Delta(\mathbf{h}) = \mathrm{Q}\left(\sqrt{2|\langle \mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h}), \mathbf{h}\rangle|^2 P}\right) - \mathrm{Q}\left(\sqrt{2|\langle \mathtt{q}_{\mathcal{B}}(\mathbf{h}), \mathbf{h}\rangle|^2 P}\right). \quad (227)$$

We consider the partition of the channel state space $\mathbb{C}^t$ via the sets

$$\begin{aligned} \mathcal{F}_1 &= \{\mathbf{h} : |h_1|^2 P \geq \beta\}, && (228)\\ \mathcal{F}_2 &= \{\mathbf{h} : |h_1|^2 P < \beta,\, |h_2|^2 P \geq \beta\}, && (229)\\ \mathcal{F}_3 &= \mathbb{C}^t - (\mathcal{F}_1 \cup \mathcal{F}_2). && (230)\end{aligned}$$

Note that

$$\mathtt{E}[\Delta(\mathbf{h})] = \sum_{i=1}^{3} \mathtt{E}[\Delta(\mathbf{h}) | \mathbf{h} \in \mathcal{F}_i] \mathtt{P}(\mathbf{h} \in \mathcal{F}_i). \quad (231)$$

By the definition of $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h})$, we have $\mathtt{E}[\Delta(\mathbf{h}) | \mathbf{h} \in \mathcal{F}_3] = 0$. On the other hand,

$$\begin{aligned} &\mathtt{E}[\Delta(\mathbf{h}) | \mathbf{h} \in \mathcal{F}_1] \mathtt{P}(\mathbf{h} \in \mathcal{F}_1)\\ &\leq \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|\langle \mathtt{q}_{\mathcal{B}}^{\mathrm{v}}(\mathbf{h}), \mathbf{h}\rangle|^2 P}\right) | \mathbf{h} \in \mathcal{F}_1\right] \mathtt{P}(\mathbf{h} \in \mathcal{F}_1) && (232)\\ &= \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|h_1|^2 P}\right) | \mathbf{h} \in \mathcal{F}_1\right] \mathtt{P}(\mathbf{h} \in \mathcal{F}_1) && (233)\\ &= \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|h_1|^2 P}\right) \mid |h_1|^2 P \geq \beta\right] \mathtt{P}(|h_1|^2 P \geq \beta) && (234)\\ &= \int_{\frac{\beta}{P}}^{\infty} \mathrm{Q}(\sqrt{2xP}) e^{-x} \mathrm{d}x && (235)\\ &\leq \frac{1}{2} \int_{\frac{\beta}{P}}^{\infty} \exp(-x(1+P)) \mathrm{d}x && (236)\\ &= \underbrace{\frac{1}{2(1+P)}}_{\leq \frac{1}{P}} \exp\left(-\beta \underbrace{\frac{1+P}{P}}_{\geq 1}\right) && (237)\\ &\leq \frac{1}{P} e^{-\beta}, && (238)\end{aligned}$$

where the second inequality is a consequence of the fact that $\mathrm{Q}(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}$. Similarly, we obtain

$$\begin{aligned} &\mathtt{E}[\Delta(\mathbf{h}) | \mathbf{h} \in \mathcal{F}_2] \mathtt{P}(\mathbf{h} \in \mathcal{F}_2)\\ &\leq \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|h_2|^2 P}\right) | \mathbf{h} \in \mathcal{F}_2\right] \mathtt{P}(\mathbf{h} \in \mathcal{F}_2) && (239)\\ &\leq \mathtt{E}\left[\mathrm{Q}\left(\sqrt{2|h_2|^2 P}\right) \mid |h_2|^2 P \geq \beta\right] \mathtt{P}(|h_2|^2 P \geq \beta) && (240)\\ &\leq \frac{1}{P} e^{-\beta}, && (241)\end{aligned}$$

where the second inequality follows since $\{\mathbf{h} : |h_2|^2 P \geq \beta\} \subset \mathcal{F}_2$, and the last inequality follows in the same manner as (238) follows from (234).

Combining the conditional expectations of $\Delta(\mathbf{h})$, we have

$$\mathtt{E}[\Delta(\mathbf{h})] \leq \frac{2}{P} e^{-\beta} = \frac{2}{P^t} e^{-g(P)}, \quad (242)$$

where we have substituted $\beta = (t-1)\log P + g(P)$ for the equality. Applying this upper bound on $\mathtt{E}[\Delta(\mathbf{h})]$ to the obvious identity $\mathtt{SER}(\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}) = \mathtt{SER}(\mathtt{q}_{\mathcal{B}}) + \mathtt{E}[\Delta(\mathbf{h})]$, we obtain (56).

We now prove (57). We have the estimates

$$\begin{aligned} \mathtt{P}(\mathbf{h} \in \mathcal{F}_2) &= \mathtt{P}(|h_1|^2 P < \beta,\, |h_2|^2 P \geq \beta) && (243)\\ &= \mathtt{P}(|h_1|^2 P < \beta) \mathtt{P}(|h_2|^2 P \geq \beta) && (244)\\ &= (1 - e^{-\frac{\beta}{P}}) e^{-\frac{\beta}{P}} && (245)\\ &\leq (1 - e^{-\frac{\beta}{P}}) && (246)\\ &\leq \frac{\beta}{P}, && (247)\end{aligned}$$

and

$$\begin{aligned} \mathtt{P}(\mathbf{h} \in \mathcal{F}_3) &= \mathtt{P}(|h_1|^2 P < \beta,\, |h_2|^2 P < \beta) && (248)\\ &= (1 - e^{-\frac{\beta}{P}})^2 && (249)\\ &\leq \frac{\beta^2}{P^2}. && (250)\end{aligned}$$

Since $g(P) \in \omega(1) \cap O(\log P)$, we have $g(P) \in O(\log P)$ in particular, and therefore, there is a constant $k > 0$ such that

$g(P) \leq k \log P$ for all sufficiently large P. For $|\mathcal{B}| \geq 2$, we can then obtain

$$\begin{aligned}
&\mathtt{R}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}}) \\
&= \underbrace{\mathtt{P}(\mathbf{h} \in \mathcal{F}_2)}_{\leq \frac{\beta}{P} \text{ by (247)}} + \underbrace{(1 + \lceil \log_2 |\mathcal{B}| \rceil)}_{\substack{\leq 2 + \log_2 |\mathcal{B}| \\ \leq 3 \log_2 |\mathcal{B}|}} \underbrace{\mathtt{P}(\mathbf{h} \in \mathcal{F}_3)}_{\leq \frac{\beta^2}{P^2} \text{ by (250)}} && (251)\\
&\leq \frac{\beta}{P} + \frac{3 \log |\mathcal{B}|}{\log 2} \frac{\beta^2}{P^2} && (252)\\
&= \frac{(t-1)\log P}{P} + \frac{g(P)}{P} + \frac{3\log|\mathcal{B}|}{\log 2} \frac{((t-1)\log P + g(P))^2}{P^2} && (253)\\
&\leq \frac{(t-1)\log P}{P} + \frac{g(P)}{P} + \frac{3(t-1+k)^2}{\log 2} \frac{\log|\mathcal{B}| \log^2 P}{P^2}. && (254)
\end{aligned}$$

This proves (57) with $\mathtt{C}_5 = \frac{3(t-1+k)^2}{\log 2}$.

Appendix J
Proof of Proposition 7

According to Proposition 3, the SNR provided by $\mathcal{B}_\delta$ is at least $\|\mathbf{h}\|^2(1-\delta)P$ for any given channel state $\mathbf{h}$. As a result,

$$\mathtt{SER}(\mathtt{q}_{\mathcal{B}_\delta}) \leq \int_0^\infty \mathrm{Q}\left(\sqrt{2x(1-\delta)P}\right) \frac{x^{t-1}e^{-x}}{\Gamma(t)} \mathrm{d}x, \quad (255)$$

where the integration variable x corresponds to a realization of the random variable $\|\mathbf{h}\|^2$. With a change of variables $u = x(1-\delta)$, we obtain

$$\mathtt{SER}(\mathtt{q}_{\mathcal{B}_\delta}) \leq (1-\delta)^{-t} \int_0^\infty \mathrm{Q}\left(\sqrt{2uP}\right) \frac{u^{t-1}\exp(-\frac{u}{1-\delta})}{\Gamma(t)} \mathrm{d}u. \quad (256)$$

It is straightforward to show that $(1-\delta)^{-t} \leq 1 + 2t\delta$ for any $\delta \leq \frac{1}{2t}$. For the integral itself, we use the bound $e^{-\frac{u}{1-\delta}} \leq e^{-u}$, which makes the integral equal to $\mathtt{SER}(\mathtt{Full})$. This concludes the proof.

Appendix K
Proof of Theorem 5

Consider a positive real-valued function $\delta(P)$ that is to be specified later on. Provided that $\delta(P)$ is sufficiently small for all sufficiently large P, we may apply Propositions 3 and 7. This gives us

$$\mathtt{SER}(\mathtt{q}_{\mathcal{B}_{\delta(P)}}) \leq \mathtt{SER}(\mathtt{Full})(1 + 2t\delta(P)), \quad (257)$$

with

$$|\mathcal{B}_{\delta(P)}| \leq \mathtt{C}_6[\delta(P)]^{-2t}. \quad (258)$$

Consider now the variable-length quantizer $\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}$. As discussed before, we may assume that $f(P) \in \omega(1) \cap O(\log P)$. Applying Proposition 5 with the choice $g(P) = \frac{f(P)}{2}$ and using the bounds in (257), (258), we obtain

$$\mathtt{R}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}) \leq \frac{(t-1)\log P}{P} + \frac{f(P)}{2P} + \frac{\mathtt{C}_5 \log(\mathtt{C}_6[\delta(P)]^{-2t}) \log^2 P}{P^2}, \quad (259)$$

with

$$\mathtt{SER}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}) \leq \mathtt{SER}(\mathtt{Full})\,(1 + 2t\delta(P)) + \frac{2}{P^t}\exp\left(-\frac{f(P)}{2}\right). \quad (260)$$

Now, let

$$\frac{\mathtt{C}_5 \log(\mathtt{C}_6[\delta(P)]^{-2t}) \log^2 P}{P} = \frac{f(P)}{2}. \quad (261)$$

We have $\delta(P) \in o(1)$ and therefore, $\delta(P)$ is positive and sufficienty small for sufficiently large P as desired. The upper bound in (260) then implies $\mathtt{d}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}) = \mathtt{d}(\mathtt{Full})$, $\mathtt{g}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}) = \mathtt{g}(\mathtt{Full})$ as $\delta(P) \in o(1)$. Combining (259) and (261), we have $\mathtt{R}(\mathtt{q}^{\mathrm{v}}_{\mathcal{B}_{\delta(P)}}) \leq (t-1)\frac{\log P}{P} + \frac{f(P)}{P}$. This concludes the proof.

Appendix L
Proof of Theorem 6

Let $\mathtt{q} = \{\mathbf{x}_n, \mathcal{E}_n, \mathtt{b}_n\}_{\mathcal{I}}$, and suppose that

$$\exists \epsilon \in (0, \tfrac{1}{2}),\, \forall P_0 \in \mathbb{R},\, \exists P \geq P_0,\, \mathtt{R}(\mathtt{q}) \leq (t-1-\epsilon)\frac{\log P}{P}. \quad (262)$$

To prove the theorem, it is sufficient to show that the strict inequality $\mathtt{d}(\mathtt{q}) < \mathtt{d}(\mathtt{Full})$ then holds.

Let $R = (t-1-\epsilon)\frac{\log P}{P}$. It is straightforward to show by contradiction that $\exists i \in \mathcal{I}$ such that $\mathtt{P}(\mathbf{h} \in \mathcal{E}_i) \geq 1 - R$. Without loss of generality, suppose that $\mathtt{P}(\mathbf{h} \in \mathcal{E}_0) \geq 1 - R$. Then, with $f(\mathbf{h})$ representing the probability density function of $\mathbf{h}$, we have

$$\begin{aligned}
\mathtt{SER}(\mathtt{q}) &= \sum_{n \in \mathcal{I}} \int_{\mathcal{E}_n} \mathrm{Q}\left(\sqrt{2|\langle \mathbf{x}_n, \mathbf{h}\rangle|^2 P}\right) f(\mathbf{h})\mathrm{d}\mathbf{h} && (263)\\
&\geq \int_{\mathcal{E}_0} \mathrm{Q}\left(\sqrt{2|\langle \mathbf{x}_0, \mathbf{h}\rangle|^2 P}\right) f(\mathbf{h})\mathrm{d}\mathbf{h} && (264)
\end{aligned}$$

According to [41, Theorem 2.1], for any $\kappa > 1$, there is a constant $\mathtt{C}_{15} > 0$ (that may depend on κ) such that $\mathrm{Q}(x) \geq \mathtt{C}_{15}\exp(-\frac{\kappa}{2}x^2)$, $\forall x \in \mathbb{R}$. Substituting this lower bound, and minimizing over all possible $\mathcal{E}_0$ and $\mathbf{x}_0$, we obtain

$$\begin{aligned}
\mathtt{SER}(\mathtt{q}) &\geq \mathtt{C}_{15} \inf_{\substack{\mathcal{E} \subset \mathbb{C}^t \\ \mathtt{P}(\mathbf{h} \in \mathcal{E}) \geq 1-R}} \inf_{\mathbf{x} \in \mathcal{X}} \int_{\mathcal{E}} \exp(-\kappa |\langle \mathbf{x}, \mathbf{h}\rangle|^2 P) f(\mathbf{h})\mathrm{d}\mathbf{h} && (265)\\
&\geq \mathtt{C}_{15} \inf_{\substack{\mathcal{E} \subset \mathbb{C}^t \\ \mathtt{P}(\mathbf{h} \in \mathcal{E}) \geq 1-R}} \int_{\mathcal{E}} \exp(-\kappa |h_1|^2 P) f(\mathbf{h})\mathrm{d}\mathbf{h}, && (266)
\end{aligned}$$

where the second inequality follows from the unitary invariance of the integral and the PDF of $\mathbf{h}$ (see (200) and the discussion that follows (200) in the proof of Theorem 4.). According to [40], the choice of $\mathcal{E}$ that minimizes (266) is of the form $\mathcal{E}' = \{\mathbf{h} : |h_1|^2 \geq r\}$, where r is a positive real number with $\mu(\mathcal{E}') = 1 - R$. In other words, one forms the solution set $\mathcal{E}'$ by starting with the points where the integrand $\exp(-\kappa|h_1|^2 P)$ takes its minimal values and then progressively adds more points until the measure of $\mathcal{E}'$ is equal to r. We have $\int_{\{\mathbf{h}:|h_1|^2 \geq r\}} f(\mathbf{h})\mathrm{d}\mathbf{h} = 1 - R$, or, equivalently $\int_r^\infty e^{-x}\mathrm{d}x = 1 - R$. Solving for r, we obtain $r = -\log(1-R)$, and thus

$$\mathtt{SER}(\mathtt{q}) \geq \mathtt{C}_{15}\int_{-\log(1-R)}^{\infty} \exp(-\kappa x P)e^{-x}\mathrm{d}x \tag{267}$$

$$= \frac{\mathtt{C}_{15}\exp[(1+\kappa P)\log(1-R)]}{1+\kappa P}. \tag{268}$$

We now have $-\log(1-R) = \sum_{n=1}^\infty R^n/n \geq R + R^2$ for sufficiently small R. Applying this lower bound, we obtain

$$\mathtt{SER}(\mathtt{q}) \geq \frac{\mathtt{C}_{15}\exp[-(1+\kappa P)(R+R^2)]}{1+\kappa P}. \tag{269}$$

Let us now choose

$$\kappa = \frac{1}{2}\left(1 + \frac{t-1}{t-1-\epsilon}\right), \tag{270}$$

and note that we have

$$\kappa \leq \frac{1}{2}\left(1 + \frac{t-1}{t-1-\frac{1}{2}}\right) \leq \frac{3}{2} \tag{271}$$

as $\epsilon \in (0, \frac{1}{2})$ by our assumption and $t \geq 2$. For the denominator of (269), we then have $1 + \kappa P \leq P + \kappa P \leq \frac{5}{2}P$ for $P \geq 1$. This gives us

$$\mathtt{SER}(\mathtt{q}) \geq \frac{2\mathtt{C}_{15}}{5P}\exp(-(R+R^2+\kappa P R^2) - \kappa P R) \tag{272}$$

Since $R = (t-1-\epsilon)\frac{\log P}{P}$ and $\kappa \leq \frac{3}{2}$, the inequality $R + R^2 + \kappa P R^2 \leq 1$ holds for all sufficiently large P. Therefore,

$$\mathtt{SER}(\mathtt{q}) \geq \frac{2\mathtt{C}_{15}}{5eP}\exp(-\kappa P R) \tag{273}$$

$$= \frac{2\mathtt{C}_{15}}{5eP}\exp\left[-\frac{1}{2}\left(1+\frac{t-1}{t-1-\epsilon}\right)(t-1-\epsilon)\log P\right] \tag{274}$$

$$= \frac{2\mathtt{C}_{15}}{5eP}\exp\left[-\left(t-1-\frac{\epsilon}{2}\right)\log P\right] \tag{275}$$

$$= \frac{2\mathtt{C}_{15}}{5e}P^{-(t-\epsilon/2)}. \tag{276}$$

This means $\mathtt{d}(\mathtt{q}) \leq t - \frac{\epsilon}{2} < t$, $\forall t \geq 2$, if $\mathtt{d}(\mathtt{q})$ ever exists. This concludes the proof.

Appendix M
Proof of (68) and (69)

We first prove (68). Suppose that $|\mathcal{S}| = 2$. We have $\rho = \frac{1}{\ell}$ and $\mathtt{CBLER}(\mathbf{x},\mathbf{h};\mathcal{S}) = \mathrm{Q}(\sqrt{\frac{1}{2}\|\mathbf{s}_1 - \mathbf{s}_2\|^2 P})$. Subject to the power constraint $\|\mathbf{s}_1\| + \|\mathbf{s}\|^2 \leq 2\ell$, the conditional BLER is minimized via binary antipodal signaling. This yields

$$\mathtt{CBLER}(\mathbf{x},\mathbf{h};\mathcal{S}) \geq \mathrm{Q}(\sqrt{2\ell|\langle\mathbf{x},\mathbf{h}\rangle|^2 P}) \tag{277}$$

$$= \mathrm{Q}(\sqrt{2\ell 2^{\frac{1}{\ell}}|\langle\mathbf{x},\mathbf{h}\rangle|^2 2^{-\rho}P}) \tag{278}$$

$$\geq \mathrm{Q}(\sqrt{4\ell|\langle\mathbf{x},\mathbf{h}\rangle|^2 2^{-\rho}P}). \tag{279}$$

Now, suppose that $|\mathcal{S}| \geq 3$. It is shown in [24, Section A] that for any $\frac{2}{3} \leq \alpha < 1$, we have

$$\mathtt{CBLER}(\mathbf{x},\mathbf{h};\mathcal{S}) \geq \frac{\alpha}{2}\mathrm{Q}\left(\sqrt{\frac{|\langle\mathbf{x},\mathbf{h}\rangle|^2 P}{2|\mathcal{S}_\alpha|}\sum_{\mathbf{s}\in\mathcal{S}_\alpha}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|^2}\right), \tag{280}$$

where $\mathcal{S}_\alpha \triangleq \{\mathbf{s}\in\mathcal{S} : \|\mathbf{s}\|^2 \leq \frac{\ell}{1-\alpha}\}$ with $|\mathcal{S}_\alpha| \geq \frac{\alpha}{2}|\mathcal{S}|$, and $\mathtt{NN}(\mathbf{s}) \in \mathcal{S}_\alpha$ represents one of the nearest neighbors of $\mathbf{s}$ in $\mathcal{S}_\alpha$. Now, let $\mathcal{B}(\mathbf{s}) = \{\mathbf{t}\in\mathbb{C}^\ell : \|\mathbf{t}-\mathbf{s}\| \leq \frac{1}{2}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|\}$ denote the open ball with radius $\frac{1}{2}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|$ centered at $\mathbf{s}$. Again, using the same ideas as in [24, Section A], it can be shown that $\{\mathcal{B}(\mathbf{s}) : \mathbf{s}\in\mathcal{S}_\alpha\}$ is a disjoint collection of open balls with $\bigcup_{\mathbf{s}\in\mathcal{S}_\alpha}\mathcal{B}(\mathbf{s}) \subset \mathcal{B}_0(2\sqrt{\frac{\ell}{1-\alpha}})$, where we denote by $\mathcal{B}_0(r)$ the open ball with radius r and center at the origin.

Now, let $\lambda(\cdot)$ be the Lebesgue measure on $\mathbb{C}^\ell$. Note that $\lambda(\mathcal{B}_0(r)) = ar^{2\ell}$ where $a = \pi^\ell/\ell!$. Since $\bigcup_{\mathbf{s}\in\mathcal{S}_\alpha}\mathcal{B}(\mathbf{s}) \subset \mathcal{B}_0(2\sqrt{\frac{\ell}{1-\alpha}})$, we have

$$\lambda\left(\bigcup_{\mathbf{s}\in\mathcal{S}_\alpha}\mathcal{B}(\mathbf{s})\right) \leq \lambda\left(\mathcal{B}_0\left(2\sqrt{\frac{\ell}{1-\alpha}}\right)\right) \tag{281}$$

$$= \frac{(4\ell)^\ell a}{(1-\alpha)^\ell}. \tag{282}$$

On the other hand, as $\{\mathcal{B}(\mathbf{s}) : \mathbf{s}\in\mathcal{S}_\alpha\}$ is a disjoint collection,

$$\lambda\left(\bigcup_{\mathbf{s}\in\mathcal{S}_\alpha}\mathcal{B}(\mathbf{s})\right) = \sum_{\mathbf{s}\in\mathcal{S}_\alpha}\lambda(\mathcal{B}(\mathbf{s})) \tag{283}$$

$$= a2^{-2\ell}\sum_{\mathbf{s}\in\mathcal{S}_\alpha}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|^{2\ell} \tag{284}$$

$$\geq a2^{-2\ell}|\mathcal{S}_\alpha|^{-(\ell-1)}\left(\sum_{\mathbf{s}\in\mathcal{S}_\alpha}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|^2\right)^\ell, \tag{285}$$

where the last inequality follows from reverse Hölder inequality. Combining this last inequality with (282), we obtain

$$\sum_{\mathbf{s}\in\mathcal{S}_\alpha}\|\mathbf{s}-\mathtt{NN}(\mathbf{s})\|^2 \leq \frac{16\ell}{1-\alpha}|\mathcal{S}_\alpha|^{\frac{\ell-1}{\ell}}. \tag{286}$$

Substituting this bound to (280), we have

$$\mathtt{CBLER}(\mathbf{x},\mathbf{h};\mathcal{S}) \geq \frac{\alpha}{2}\mathrm{Q}\left(\sqrt{\frac{8\ell|\langle\mathbf{x},\mathbf{h}\rangle|^2 P}{(1-\alpha)|\mathcal{S}_\alpha|^{1/\ell}}}\right) \tag{287}$$

$$\geq \frac{\alpha}{2}\mathrm{Q}\left(\sqrt{\frac{8\ell|\langle\mathbf{x},\mathbf{h}\rangle|^2 P}{(1-\alpha)|\mathcal{S}|^{1/\ell}(\alpha/2)^{1/\ell}}}\right) \tag{288}$$

$$\geq \frac{1}{3}\mathrm{Q}\left(\sqrt{72\ell|\langle\mathbf{x},\mathbf{h}\rangle|^2 2^{-\rho}P}\right), \tag{289}$$

where the second inequality follows since $|\mathcal{S}_\alpha| \geq \frac{\alpha}{2}|\mathcal{S}|$, and the last inequality follows from the substitutions $\alpha = \frac{2}{3}$, $|\mathcal{S}| = 2^{\rho\ell}$ and the fact that $(\frac{1}{3})^{1/\ell} \geq \frac{1}{3}$, $\ell \geq 1$. Combining (289) and (279) proves (68).

We now prove (69). In order to handle the cases where $\rho \leq 2$ or $\rho \notin 2\mathbb{Z}$, we consider codes with a (possibly) slightly larger rate by letting $\bar{\rho} = 2\lfloor\rho/2 + 1\rfloor$. Now, for $\ell = 1$, we let $\mathcal{S}_0 = \{-1 + i2^{1-\bar{\rho}/2} : i = 0, \ldots, 2^{\bar{\rho}/2} - 1\}^2$ be a QAM-alphabet with rate $\bar{\rho} \geq \rho$. For each signal point $s \in \mathcal{S}_0$, there are at most 4 nearest neighbors. Moreover, each nearest neighbor of s is $2^{1-\bar{\rho}/2}$-far from s. Since $\bar{\rho} \leq \rho + 2$, each nearest neighbor of s is at least $2^{2-\rho/2}$-far from s. Hence, $\mathtt{CBLER}(\mathbf{x},\mathbf{h};\mathcal{S}_0) \leq 4\mathrm{Q}\left(\sqrt{8|\langle\mathbf{x},\mathbf{h}\rangle|^2 2^{-\rho}P}\right)$. For a general $\ell \geq 2$, we set $\mathcal{S}_0$ to be the ℓth Cartesian power of the same QAM alphabet. A simple union bound then leads to (69).

Appendix N
Proof of Theorem 7

For achievability, we proceed in the same manner as we have done in Section VIII-B for the binary modulation case. Consider a fixed beamforming codebook $\mathcal{B}$. We synthesize the variable-length quantizer $\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}$ according to the same rules discussed in Section VIII-B. The only difference is that for a given $g(P) \in \omega(1) \cap O(\log P)$, we choose

$$\beta = \frac{2}{\mathtt{C}_8}\left[(t-1)(1-r)\log P + g(P)\right]2^{\rho} \tag{290}$$

instead of $\beta = (t-1)\log P + g(P)$. Then, using (70) and the same arguments as in the proof of Proposition 5, we can obtain

$$\mathtt{BLER}(\mathtt{q}_{\mathcal{B}}^{\mathrm{v}};\mathcal{S}) \leq \mathtt{BLER}(\mathtt{q}_{\mathcal{B}};\mathcal{S}) + \frac{2}{P^{t(1-r)}}e^{-g(P)}, \tag{291}$$

and

$$\mathtt{R}(\mathtt{q}_{\mathcal{B}}^{\mathrm{v}}) \leq 1 + \mathtt{C}_{16}\left(\frac{2^{\rho}\log P}{P} + \frac{2^{\rho}g(P)}{P} + \frac{2^{2\rho}\log|\mathcal{B}|\log^2 P}{P^2}\right), \tag{292}$$

where $\mathtt{C}_{16} > 0$ is independent of ρ and P. Now, following the same steps as in Sections VIII-C and VIII-D, we can obtain the achievability result in (73). On the other hand, using the lower bound in (68) and same ideas as in the proof of Theorem 6 in Appendix L, we obtain the converse result in (74). We omit the details as they are straightforward.

References

[1] D. J. Love, R. W. Heath, Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[2] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[3] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the power loss when transmit beamforming relies on finite-rate feedback," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1948–1957, Jul. 2005.

[4] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1298–1310, Sept. 2006.

[5] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[6] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[7] ——, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback", *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[8] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

[9] J. Zheng, E. R. Duni, and B. D. Rao, "Analysis of multiple-antenna systems with finite-rate feedback using high-resolution quantization theory," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1461–1476, Apr. 2007.

[10] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[11] S. Ekbatani and H. Jafarkhani, "Combining beamforming and space-time coding using quantized feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 898–908, Mar. 2008.

[12] ——, "Combining beamforming and space-time coding using noisy quantized feedback," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1280–1286, May 2009.

[13] M. Xu, D. Guo, and M. L. Honig, "Multicarrier beamforming with limited feedback: A rate distortion approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 916–927, Feb. 2013.

[14] M. Agarwal, D. Guo, and M. L. Honig, "Limited-rate channel state feedback for multicarrier block fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6116–6132, Dec. 2010.

[15] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," *IEEE Military Commun. Conf. (MILCOM)*, Oct. 2003.

[16] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

[17] W. Santipach and K. Mamat, "Tree-Structured Random Vector Quantization for Limited-Feedback Wireless Channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3012–3019, Sept. 2011.

[18] C. K. Au-Yeung, D. J. Love, and S. Sanayei, "Trellis coded line packing: Large dimensional beamforming vector quantization and feedback transmission," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1844–1853, June 2011.

[19] S. A. Jafar and S. Srinivasa, "On the optimality of beamforming with quantized feedback," *IEEE Trans. Commun.*, nol. 55, no. 12, pp. 2288–2302, Dec. 2007.

[20] E. Koyuncu, Y. Jing, and H. Jafarkhani, "Distributed beamforming in wireless relay networks with quantized feedback," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1429–1439, Oct. 2008

[21] Y. Zhao, R. Adve, and T. J. Lim, "Beamforming with limited feedback in amplify-and-forward cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5145–5149, Dec. 2008

[22] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed transmit beamforming using feedback control," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 411–426, Jan. 2010.

[23] E. Koyuncu and H. Jafarkhani, "Distributed beamforming in wireless multiuser relay-interference networks with quantized feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4538–4576, July 2012.

[24] ——, "On the structure of limited-feedback beamforming codebooks for amplify-and-forward relay networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2874–2895, May 2012.

[25] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1959–1975, May. 2009.

[26] J. Choi, B. Clerckx, N. Lee, and G. Kim, "A new design of polar-cap differential codebook for temporally/spatially correlated MISO channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 703–711, Feb. 2012.

[27] Y. Huang, L. Yang, M. Bengtsson, and B. Ottersten, "Exploiting long-term channel correlation in limited feedback SDMA through channel phase codebook," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1217–1228, Mar. 2011.

[28] K. K. Mukkavilli, A. Sabharwal, and E. Erkip, "On beamforming with finite-rate feedback for multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[29] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[30] Y. Ma, D. Zhang, A. Leith, and Z. Wang, "Error performance of transmit beamforming with delayed and limited feedback," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1164–1170, Mar. 2009.

[31] T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1654–1669, Apr. 2011.

[32] D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig, "QAM and PSK codebooks for limited feedback MIMO beamforming," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1184–1196, Apr. 2009.

[33] R. Bhagavatula and R. W. Heath, Jr., "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800–811, Feb. 2011.

[34] V. Pieterse and P. E. Black, Eds., *Distionary of Algorithms and Data Structures*, U.S. National Institute of Standards and Technology. [Online]. Available: http://www.nist.gov/dads/

[35] R. Porat, E. Ojard, N. Jindal, M. Fischer, and V. Erceg, "Improved MU-MIMO performance for future 802.11 systems using differential feedback," *IEEE Inf. Theory and Appl. Conf. (ITA)*, Feb. 2013.

[36] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[37] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[38] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.
[39] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 37, no. 1, pp. 31–42, Jan. 1989.
[40] P. K. C. Wang, "On a class of optimization problems involving domain variations," *Intl. Symp. on New Trends in System Analysis*, Dec. 1976.
[41] F. D. Côté, I. N. Psaromiligkos, and W. J. Gross, "A Chernoff-type lower bound for the Gaussian Q-function," Mar. 2012. [Online]. Available: http://arxiv.org/pdf/1202.6483