Title

Delay-Limited and Ergodic Capacities of MIMO Channels With Limited Feedback

Journal

IEEE Transactions on Communications, 64(9)

ISSN

0090-6778

Authors

Koyuncu, Erdem
Jafarkhani, Hamid

Publication Date

2016

DOI

10.1109/tcomm.2016.2590546

Peer reviewed

Delay-Limited and Ergodic Capacities of MIMO Channels with Limited Feedback

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

Abstract

We consider a fixed data rate slow-fading MIMO channel with a long-term power constraint P at the transmitter. A relevant performance limit is the delay-limited capacity, which is the largest data rate at which the outage probability is zero. It is well-known that if both the transmitter and the receiver have full channel state information (CSI) and if either of them has multiple antennas, the delay-limited capacity is non-zero and grows logarithmically with P. Achieving even a positive delay-limited capacity however becomes a difficult task when the CSI at the transmitter (CSIT) is imperfect. In this context, the standard partial CSIT model where the transmitter has a fixed finite bits of quantized CSI feedback for each channel state results in zero delay-limited capacity. We show that by using a variable-length feedback scheme that utilizes different number of feedback bits for different channel states, a non-zero delay-limited capacity can be achieved if the feedback rate is greater than 1 bit per channel state. Moreover, we show that the delay-limited capacity loss due to finite-rate feedback decays at least inverse linearly with respect to the feedback rate. We also discuss applications to ergodic MIMO channels.

I. Introduction

Minimizing the outage probabilities of MIMO systems with different assumptions on transmitter/receiver channel knowledge and power constraints is a classical fundamental problem of communication theory and has been the subject of many publications. In particular, it is known [1], [2] that transmit precoding with power control provides the best-possible

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

The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Honolulu, Hawaii, USA, June 29–July 4, 2014.

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.

outage probabilities of a fixed-rate $t \times r$ MIMO system. A remarkable result of [1] is that unless $t = r = 1$, then for any given long-term power constraint $P > 0$, transmission with zero-outage probability is possible for all rates up to a positive delay-limited capacity $C(P) > 0$. Moreover, the capacity $C(P)$ grows with a multiplexing gain of $m \triangleq \min\{t, r\}$, i.e., $C(P) = m \log P + O(1)$ as $P \to \infty$.

The results of [1] rely on the assumption that both the transmitter and the receiver knows the channel state information (CSI) perfectly. In practice, the receiver can indeed acquire the CSI via training sequences from the transmitter. Obtaining CSI at the transmitter (CSIT) is however more difficult and requires feedback from the receiver. One complication in this context is that the MIMO CSI is, in general, an arbitrary complex-valued matrix, and therefore feeding it back unquantized would require "an infinite rate of feedback." This motivates the current paper where we study the delay-limited capacities of MIMO systems with partial quantized CSIT. The core element of our partial CSIT model is a channel quantizer that specifies for every channel state (i) the sequence of feedback bits to be fed back from the receiver to the transmitter, and (ii) for each such sequence, the precoding matrix to be used by the transmitter. We design optimal quantizers that maximize the delay-limited capacity subject to the long-term power constraint P of the transmitter and the rate constraint R of the feedback link measured in bits per fading block.

A. Related Work

There has been a lot of work on power control with rate-limited feedback. In particular, distributed quantized power control algorithms for single-antenna cellular systems have been studied in [3], [4]. Also, [5] has considered the quantized power control problem for single-antenna CDMA systems. The achievable gains with limited-feedback power control have been studied in [6], [7] for MISO fading channels. In the more general MIMO setting, [8], [9] have determined the achievable diversity-multiplexing gains with quantized power control. The limited-feedback power control problem has also been studied for relay networks [10].

Most of the previous work on limited feedback MIMO power control [6]–[9] have focused on a particular class of quantizers that we shall here refer to as "fixed-length quantizers (FLQs)," in which the number of feedback bits is a fixed natural number for every channel state. According to these studies, for a fixed target data rate, the best possible outage probability with a rate-R FLQ decays as $O(P^{-d})$ where $d = \sum_{i=1}^{2^R}(rt)^i$ is the outage exponent of the system and r is the target multiplexing gain. Compared to the case of a no-CSIT system [11], FLQs can therefore significantly improve the outage exponent of the system. On the other hand, unlike the case of a full-CSIT system, FLQs fail to achieve zero outage probability for any finite P and R. Hence, the delay-limited capacity of the MIMO channel with finite-rate FLQs is zero. Moreover, for any R, the outage exponents of FLQs will decrease from d to 0 as the target multiplexing gain r increases from 0 to m [8]. In particular, transmitting within a constant gap of the full-CSIT delay-limited capacity (i.e. transmitting with a multiplexing gain of m) will result in constant outage probability even as $P \to \infty$.

Outage probability and delay-limited capacity are relevant performance measures when the channel codewords are assumed to span only a fixed finite number of, say B, fading blocks. This is also commonly referred to as the "B-block fading" scenario [1]. If coding can be performed over an arbitrarily large number of blocks ($B \to \infty$), the relevant performance measure is the ergodic capacity [12]. Ergodic capacities of MIMO channels with limited feedback have previously been studied with different approaches based on Grassmannian line packings [13]–[15], vector quantizer design algorithms [16]–[20], high resolution approximations [21], [22], random vector quantizers [23]–[28], and several other techniques [29]–[34]. The performance gains of limited feedback in the context of different performance measures, such as bit-error or symbol-error probabilities, have also been analyzed [35]–[37]. Also, achievable rates in MIMO broadcast channels with limited feedback have been determined [38], [39]. In particular, using random quantizers, it is shown in [27] that for a general MIMO system with a short-term power constraint, the ergodic capacity loss due to quantization decays at least exponentially with the feedback rate.

B. Summary of Main Results

The previous studies (as outlined in Section I-A) leave one with the impression that outage is unavoidable when the transmitter CSI knowledge is less than perfect, or when the receiver feedback rate is finite. In other words, the delay-limited capacities with limited feedback are zero regardless of how high the rate of feedback is. We show that the unachievability of a non-zero delay-limited capacity in the existing work is due to the implicit assumption of FLQs for feedback information. In fact, by using variable-length quantizers (VLQs) [40]–

TABLE I: Capacity Losses with FLQs and VLQs Relative to a Full-CSIT System

	FLQs	VLQs
B-block delay-limited capacity (short-term power constraint)	Capacity is 0.	Capacity is 0.
B-block delay-limited capacity (long-term power constraint)	Capacity is 0.	$O(1/R)$
Ergodic capacity (short-term power constraint)	$O(\exp(-c_1 R))$	$O(\exp(-c_1 R))$
Ergodic capacity (long-term power constraint)	$O(\exp(-c_2 R))$	$O(\exp(-c_2 R))$

[45], which allow different number of feedback bits for different channel states, we show that a positive delay-limited capacity can be achieved for any feedback rate greater than 1 bit per channel state. Moreover, we show that the capacity with VLQs can be made arbitrarily close to the capacity with perfect CSIT by allowing a sufficiently large (but finite) feedback rate. In fact, our results demonstrate that the delay-limited capacity loss due to quantization decays at least inverse linearly with the feedback rate. We also discuss generalizations to multi-user MIMO multicast networks.

Analyzing general-rank quantized precoding systems is known to be much more challenging than analyzing the simpler rank-1 quantized beamforming systems [27], [33], [34]. Another contribution of this paper is a new set of versatile technical tools that are applicable to a variety of quantized MIMO precoding problems. As applications of these tools (in addition to the results above), we also show that the ergodic capacity loss due to quantization decays exponentially with the feedback rate for both short-term and long-term power constraints. The short-term case has previously been shown in [27]; we provide here a much simpler proof that utilizes FLQs with an explicit construction as opposed to random FLQs.

In the light of the above main results, bounds on the capacity losses with FLQs and VLQs for different capacity measures are shown in Table I as a function of the feedback rate R. The bounds hold for every sufficiently large P, and c_1, c_2 are constants that are independent of R and P.

C. Organization of the Paper

The rest of the paper is organized as follows: In Section II, we introduce the system model, the formal definitions of the channel quantizers and the corresponding delay-limited capacities. In Section III, we study the limited-feedback delay-limited capacities with scaled-identity covariance matrices. In Section IV, we consider the case of arbitrary covariance matrices. In Section V, we extend our results to general B-block fading channels for $B > 1$. In Section VI, we provide applications to ergodic channels with long-term or short-term power constraints. Finally, in Section VII, we draw our main conclusions. Some of the technical proofs are provided in the appendices.

D. Notation

The sets $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{Z}$ are the sets of complex numbers, real numbers, and integers, respectively. For any given $z \in \mathbb{C}$, $\Re(z)$ and $\Im(z)$ represent the real and imaginary parts of z, respectively, and $|z|$ is the norm of z. For any set A, we denote by A^m the set of all $m \times 1$ vectors whose components are members of A. Similarly, $A^{m \times n}$ is the set of all $m \times n$ matrices whose entries are members of A. The set $A_{>0}$ is the set of positive elements of A, and $A_{\geq 0}$ is defined similarly. We use the notation $[h_{ij}]_{r \times t}$ to represent a matrix with r rows, t columns, and the entry h_{ij} in its ith row, jth column. For any given matrix $\mathbf{A}$, the quantity $\|\mathbf{A}\|$ is the Frobenius norm of $\mathbf{A}$. We denote by $\mathbf{I}_t$ the $t \times t$ identity matrix. We let $\mathcal{CN}(\mathbf{Q})$ represent a circularly-symmetric complex Gaussian random variable with covariance matrix $\mathbf{Q}$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{Q})$ means that $\mathbf{n}$ is distributed as $\mathcal{CN}(\mathbf{Q})$. Finally, $\mathtt{P}(\cdot)$ and $\mathtt{E}[\cdot]$ represent the probability and the expected value, respectively.

II. Preliminaries

We begin our discussion with the simple case of MIMO systems with scaled-identity precoding matrices. We also assume that each channel codeword spans only one fading block. We introduce the system model, and the notions of delay-limited capacities and quantizers. We later generalize these definitions to MIMO systems with arbitrary precoding matrices, and then to general B-block MIMO fading channels.

A. System Model and Performance Measures

Consider a MIMO system with t transmitter antennas and r receiver antennas. Denote the channel from the transmitter antenna j to the receiver antenna i by h_{ij}, and let $\mathbf{H} = [h_{ij}]_{r\times t}$ represent the entire channel state. We assume that the components of $\mathbf{H}$ are independent and identically distributed as $\mathcal{CN}(1)$. The transmitted symbol $\mathbf{s} \in \mathbb{C}$ and the received vector $y \in \mathbb{C}^{r\times 1}$ have the input-output relationship $\mathbf{y} = \mathbf{H}\mathbf{s}+\mathbf{n}$, where the noise vector $\mathbf{n} \sim \mathcal{CN}(\mathbf{I}_r)$ is independent of $\mathbf{h}$. We assume that the receiver knows $\mathbf{H}$ perfectly.

Let $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_m]$ denote the unordered positive eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ or $\mathbf{H}\mathbf{H}^\dagger$. For a fixed $\mathbf{H}$, suppose that the input symbol $\mathbf{s}$ is chosen as $\mathbf{s} \sim \mathcal{CN}(\frac{\sigma}{t}\mathbf{I}_t)$ for some $\sigma \geq 0$. The channel capacity with this strategy is $c(\sigma, \boldsymbol{\lambda}) \triangleq \sum_{i=1}^{m} \log(1 + \frac{\sigma}{t}\lambda_i)$ nats/sec/Hz. The transmission power (averaged over the Gaussian symbol $\mathbf{s}$) is $\mathtt{E}[\|\mathbf{s}\|^2] = \sigma$. For a given target data transmission rate $c \geq 0$, an outage event occurs if $c(\sigma, \boldsymbol{\lambda}) < c$.

When $\mathbf{H}$ is random, we define the outage probability as the fraction of channel realizations for which outage events occur. To be more precise, consider an arbitrary power-control mapping $\sigma : \mathbb{R}_{>0}^m \to \mathbb{R}_{\geq 0}$, and suppose that the transmitter transmits with power $\sigma(\boldsymbol{\lambda})$ given that the channel eigenvalues are given by $\boldsymbol{\lambda}$. Then, for a given target data transmission rate $c \geq 0$, we define the outage probability with mapping σ as $\mathtt{P}_{\mathtt{out}}(\sigma; c) \triangleq \mathtt{P}(c(\sigma(\boldsymbol{\lambda}), \boldsymbol{\lambda}) < c)$. The average transmission power consumed by the mapping σ is given by $\mathcal{P}(\sigma) \triangleq \mathtt{E}[\sigma(\boldsymbol{\lambda})]$, where the expectation is over all possible eigenvalues.

In this paper, our main focus is on mappings that can achieve zero outage probability. Correspondingly, we define the delay-limited (zero-outage) capacity $c(\sigma) \triangleq \sup\{c \geq 0 : \mathtt{P}_{\mathtt{out}}(\sigma; c) = 0\}$ with σ as the supremum of all transmission rates where the outage probability with σ is zero. With perfect CSIT, an optimal "full-CSIT mapping" $\sigma_\star \triangleq \arg\max\{c(\sigma) : \mathcal{P}(\sigma) \leq P\}$ exists [1] and provides the maximum achievable capacity (with scaled-identity covariance matrices) $c(P) \triangleq c(\sigma_\star)$ of the MIMO channel subject to a long-term power constraint $P \geq 0$ at the transmitter. An explicit form of $\sigma_\star$ is given by $\sigma_\star(\boldsymbol{\lambda}) = \min\{\sigma \in \mathbb{R}_{\geq 0} : c(\sigma, \boldsymbol{\lambda}) = c\}$, where c satisfies $\mathtt{E}[\sigma_\star(\boldsymbol{\lambda})] = P$. In other words, for each realization of the eigenvalues, one spends just enough power to avoid outage for a target data rate that should be chosen to satisfy the long-term power constraint of the transmitter.

In particular, for $t = 1$ and $r > 1$, we have the full-CSIT mapping $\sigma_\star(\lambda_1) = \frac{(r-1)P}{\lambda_1}$ for a SIMO system with $c(\sigma_\star) = \log(1 + (r-1)P)$. For $r = 1$ and $t > 1$ (a MISO system), we can calculate $\sigma_\star(\lambda_1) = \frac{(t-1)P}{\lambda_1}$ with $c(\sigma_\star) = \log(1+\frac{t-1}{t}P)$. For $r = t = 1$ (a SISO system), $c(\sigma_\star) = 0, \forall P \geq 0$ (Supporting any given positive data rate with zero outage probability requires infinite average power.). In general, for $t > 1$ and $r > 1$, no simple closed-form expressions are known for the full-CSIT mappings and the corresponding capacities. However, asymptotic expressions are available for the $P \to \infty$ regime. Specifically, for $m \triangleq \min\{t,r\}$ and $\overline{m} \triangleq \max\{t,r\}$, it can be shown (see, e.g., [1] and [46]) that

$$c(P) = m \log P - m \log\left(t \prod_{i=1}^{m} \frac{\Gamma(\overline{m} - i + 1 - \frac{1}{m})}{\Gamma(\overline{m} - i + 1)}\right) + o(1),\ \overline{m} > 1. \tag{1}$$

B. Channel Quantizers for Partial-CSIT Systems

The capacity-maximizing full-CSIT mapping $\sigma_\star$ requires perfect knowledge of $\sigma_\star(\boldsymbol{\lambda})$ at the transmitter. In practice, for each channel state, $\sigma_\star(\boldsymbol{\lambda})$ should then be fed back from the receiver to the transmitter. Since $\sigma_\star(\boldsymbol{\lambda})$ can assume an arbitrary positive real number, feeding it back unquantized would require an infinite rate of feedback. We thus consider a partial-CSIT scenario where the transmitter transmits with quantized power levels according to the receiver's finite-rate feedback information. Such a system can be modeled by a "channel quantizer" as we describe in the following.

Let $\mathcal{I}$ be a countable index set. We use the notation $\{a_n\}_{n\in\mathcal{I}}$ for the set $\{a_n : n \in \mathcal{I}\}$. For a given index set $\mathcal{I}$, let $\{\sigma_n\}_{n\in\mathcal{I}}$ be a set of quantized transmission power levels with $\{\sigma_n\}_{n\in\mathcal{I}} \subset \mathbb{R}_{\geq 0}$. Also, let $\{\mathcal{E}_n\}_{n\in\mathcal{I}}$ be a collection of mutually disjoint subsets of $\mathbb{R}^m_{>0}$ with $\cup_{n\in\mathcal{I}}\mathcal{E}_n = \mathbb{R}^m_{>0}$. Finally, let $\{\mathtt{b}_n\}_{n\in\mathcal{I}} \subset \{0,1\}^+$ be a collection of feedback binary codewords, where $\{0,1\}^+ \triangleq \{\mathtt{0},\mathtt{1},\mathtt{00},\mathtt{01},\ldots\}$ is the set of all non-empty binary codewords. We assume the collection $\{\mathtt{b}_n\}_{n\in\mathcal{I}}$ forms a prefix-free code. We call the collection of triples $\mathtt{q} \triangleq \{[\sigma_n,\,\mathcal{E}_n,\,\mathtt{b}_n]\}_{n\in\mathcal{I}}$ a quantizer $\mathtt{q}$. We call $\mathtt{q}$ an infinite-level quantizer if $\mathcal{I}$ is an infinite set. Otherwise, we call $\mathtt{q}$ a finite-level (or, to be more specific, an $|\mathcal{I}|$-level) quantizer.

The quantizer $\mathtt{q}$ corresponds to 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 $\boldsymbol{\lambda} \in \mathcal{E}_n$. Such an index n always exists and is unique as $\{\mathcal{E}_n\}_{n\in\mathcal{I}}$ forms a disjoint covering of $\mathbb{R}^m_{>0}$. The transmitter recovers the index n and transmits with power σ_n. The recovery of n by the transmitter is always possible since $\mathtt{b}_n$'s are distinct. We write $\mathtt{q}(\boldsymbol{\lambda}) = \sigma_n$ whenever $\boldsymbol{\lambda} \in \mathcal{E}_n$ to emphasize the quantization operation. We call the set $\{\sigma_n\}_{n\in\mathcal{I}}$ the quantizer codebook.

For any $\mathtt{b} \in \{0,1\}^+$, let $\ell(\mathtt{b})$ denote the "length" of $\mathtt{b}$. For example, $\ell(\mathtt{0}) = 1$, $\ell(\mathtt{01}) = 2$. A quantizer $\mathtt{q}$ is called a fixed-length quantizer (FLQ) if $\ell(\mathtt{b}_i) = \ell(\mathtt{b}_j)$ for every $i, j \in \mathcal{I}$. Otherwise, we call $\mathtt{q}$ a variable-length quantizer (VLQ). In either case, the quantization rate $\mathcal{R}(\mathtt{q}) \triangleq \sum_{n \in \mathcal{I}} \mathsf{P}(\boldsymbol{\lambda} \in \mathcal{E}_n)\ell(\mathtt{b}_n)$ measures the rate of feedback information from the receiver to the transmitter. Also, since quantizers are special cases of the mappings discussed in Section II-A, we let $c(\mathtt{q})$ denote the capacity with a quantizer $\mathtt{q}$.

C. Delay-Limited Capacities with FLQs and VLQs

Our goal is to design quantizers that maximize the delay-limited capacity of the channel subject to the rate constraint of the feedback link and the power constraint of the transmitter. Namely, we are interested in determining the capacities $c_R(P) \triangleq \sup\{c(\mathtt{q}) : \mathcal{P}(\mathtt{q}) \leq P, \mathcal{R}(\mathtt{q}) \leq R\}$ and the quantizers that can achieve them. We also compare the capacities $c_R(P)$ with the capacity $c(P) = c(\sigma_\star)$ with perfect CSIT.

The capacity with full-CSIT grows logarithmically with P and provides a multiplexing gain of m. In general, in order to achieve even a positive capacity, the transmitter should be able to transmit with arbitrarily large powers so that outage events can be avoided even at extremely ill-conditioned channel states. An important byproduct of this observation is that a finite-rate FLQ (or, in general, a finite-level quantizer) cannot provide a positive capacity at any P as an FLQ can support only a finite number of transmission powers. VLQs appear as natural candidates in this context as a VLQ can utilize infinitely many, arbitrarily large transmission powers while still retaining a finite feedback rate and a finite average power.

In the following, we show that VLQs can in fact achieve a positive delay-limited capacity at any feedback rate and power constraint, i.e., for every $P > 0$ and $R > 1$, we have $c_R(P) > 0$. Moreover, we construct a family of VLQs whose capacities converge to the full-CSIT capacity uniformly as the feedback rate grows to infinity. In other words, we show that as $R \to \infty$, we have $c_R(P) \to c(P)$ for every $P \geq 0$. Later, we will show how to obtain analogous results for general precoding matrices.

III. VLQs for Scaled Identity Precoding Matrices

A first natural idea for a VLQ design might be to directly quantize the optimal transmission power $\sigma_\star(\boldsymbol{\lambda})$ using a scalar quantizer. However, the performance of quantizers designed in this manner are very difficult to analyze as there is no simple expression for the PDF of $\sigma_\star(\boldsymbol{\lambda})$. We thus take an indirect route by designing a vector quantizer for $\boldsymbol{\lambda}$.

We begin by defining a set of auxiliary binary codewords. For $n \in \mathbb{Z}_{\geq 0}$, let $\mathtt{a}_n = \mathtt{0}\mathtt{0}^n\mathtt{1}$, and for $n \in \mathbb{Z}_{<0}$, let $\mathtt{a}_n = \mathtt{1}\mathtt{0}^{|n|-1}\mathtt{1}$. The set of auxiliary binary codewords $\{\mathtt{a}_n\}_{n\in\mathbb{Z}} = \{\ldots, \mathtt{1001}, \mathtt{101}, \mathtt{11}, \mathtt{01}, \mathtt{001}, \mathtt{0001}, \ldots\}$ then defines a prefix-free code. For a given arbitrary real number $\beta > 1$, let

$$\mathcal{E}_{\{-1\}^m} \triangleq [\tfrac{1}{\beta}, \beta)^m \tag{2}$$

$$\mathtt{b}_{\{-1\}^m} \triangleq \mathtt{0}, \tag{3}$$

$$\sigma_{\{-1\}^m} \triangleq \sigma_\star([\tfrac{1}{\beta} \ \cdots \ \tfrac{1}{\beta}]). \tag{4}$$

Also, for any $[i_1 \cdots i_m] \in \mathbb{Z}^m \backslash \{-1, 0\}^m$, we define

$$\mathcal{E}_{[i_1\cdots i_m]} \triangleq [\beta^{i_1}, \beta^{i_1+1}) \times \cdots \times [\beta^{i_m}, \beta^{i_m+1}) \tag{5}$$

$$\mathtt{b}_{[i_1\cdots i_m]} \triangleq \mathtt{1}\mathtt{a}_{i_1} \cdots \mathtt{a}_{i_m}, \tag{6}$$

$$\sigma_{[i_1\cdots i_m]} \triangleq \sigma_\star([\beta^{i_1} \ \cdots \ \beta^{i_m}]). \tag{7}$$

Now, let $\mathcal{K} \triangleq \{\{-1\}^m\} \cup (\mathbb{Z}^m \backslash \{-1, 0\}^m)$. We consider the quantizer

$$\mathtt{q}_\beta \triangleq \{\sigma_\kappa, \mathcal{E}_\kappa, \mathtt{b}_\kappa\}_{\kappa\in\mathcal{K}}. \tag{8}$$

For each $\beta > 1$, we have a different quantizer with different encoding cells and power levels. As we shall explain in the following, the purpose of the parameter β is to control the transmission power and the rate of the quantizer $\mathtt{q}_\beta$.

In general, $\mathtt{q}_\beta$ quantizes the unordered eigenvalues $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_m]$ to a quantized power level $\mathtt{q}_\beta(\boldsymbol{\lambda})$. To gain intuition on the structure of $\mathtt{q}_\beta$, we first discuss the special case $m = 1$ as illustrated in Fig. 1. The horizontal axis represents λ_1 and the vertical axis represents the instantaneous transmission power (with $\mathtt{q}_\beta$ or $\sigma_\star$) given λ_1. Both axes are in the logarithmic scale. The thick black line represents the full-CSIT mapping $\sigma_\star(\lambda_1) = \frac{\alpha}{\lambda_1}$, where $\alpha \triangleq (r-1)P$, and the piecewise gray line represents the quantizer $\mathtt{q}_\beta$.

First, we observe that for every $\beta > 1$ and any given λ_1, the transmission power with $\mathtt{q}_\beta$ is no less than the transmission power with the full-CSIT mapping $\sigma_\star$, i.e., for all $\beta > 1$ and $\lambda_1 > 0$, we have $\mathtt{q}_\beta(\lambda_1) \geq \sigma_\star(\lambda_1)$. Hence, the quantizer $\mathtt{q}_\beta$ can avoid outage whenever $\sigma_\star$ does, and therefore, we have $c(\mathtt{q}_\beta) \geq c(\sigma_\star)$.

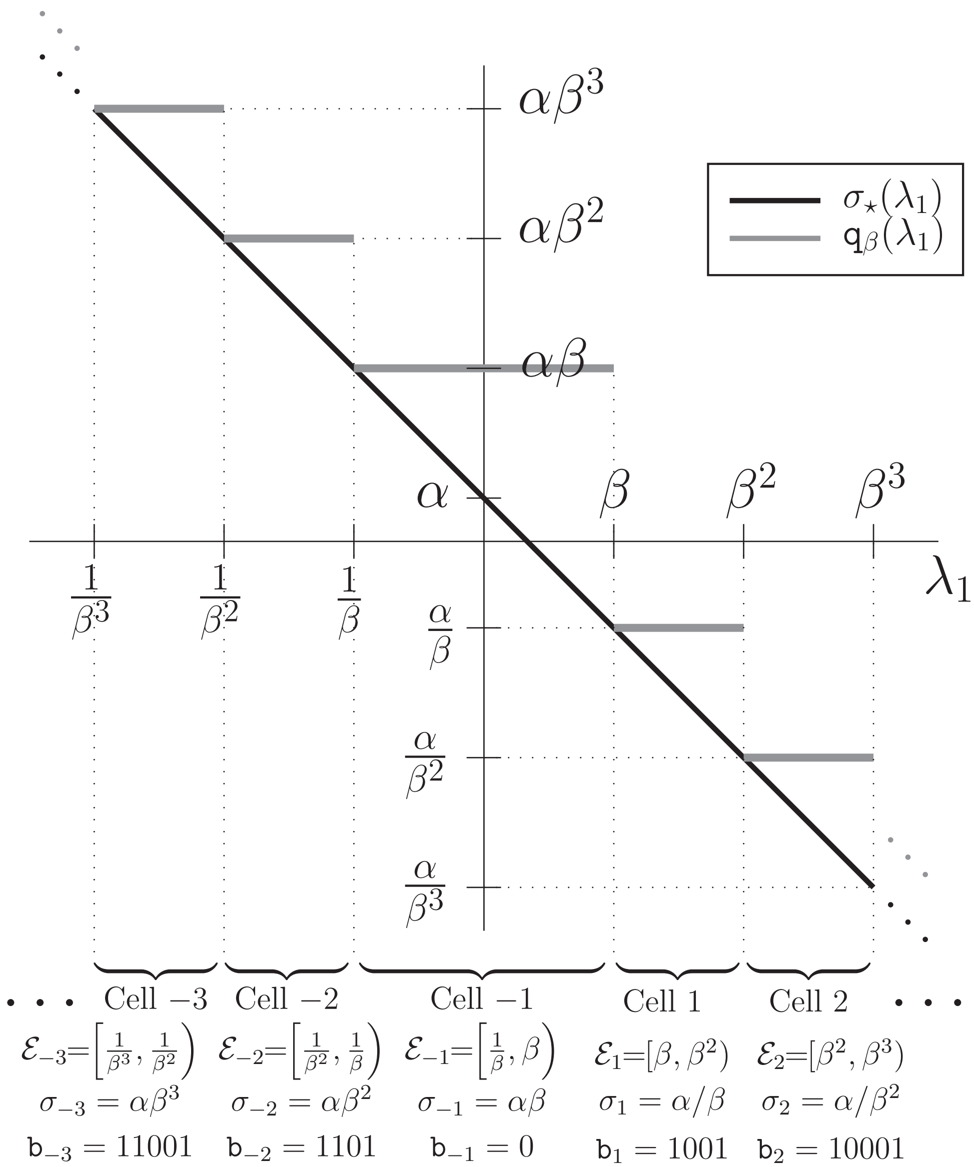

Fig. 1: The quantizer $\mathtt{q}_\beta$ for the special case $m = 1$.

Second, we observe that the quantizer $\mathtt{q}_\beta$ is a "log-uniform deadzone quantizer" meaning that the lengths of all its cells (except the cell $\mathcal{E}_{-1}$) in the logarithmic domain is the same. One motivation for choosing a log-uniform structure is that it leads to quantizers with bounded transmission powers. Indeed, we have $\mathtt{q}_\beta(\lambda) \leq \beta^2\sigma_\star(\lambda)$ for every $\lambda > 0$, and therefore, $\mathcal{P}(\mathtt{q}_\beta) \leq \beta^2\mathcal{P}(\sigma_\star)$ (To see why the first inequality is true, note that multiplying $\sigma_\star(\gamma)$ by β^2 is equivalent to shifting the thick black line in Fig. 1 above by $2\log\beta$ units. This shifted version of $\sigma_\star$ serves as an upper bound on $\mathtt{q}_\beta(\lambda)$ for every λ, which proves the inequality.).

Fortunately, the log-uniform structure of $\mathtt{q}_\beta$ that guarantees its bounded power consumption also guarantees a bounded feedback rate. Indeed, for $n \to -\infty$, the lengths of the quantization cells $\mathcal{E}_n$ decay exponentially with $|n|$ as $O(\beta^{-|n|})$. This results in $\mathtt{P}(\lambda \in \mathcal{E}_n) \in O(\beta^{-|n|})$ as the PDF $f(\lambda_1) = e^{-\lambda_1}\lambda_1^{\overline{m}-1}/\Gamma(\overline{m})$, $\lambda_1 \geq 0$ of λ_1 is bounded from above. On the other hand, as $n \to \infty$, the PDF of λ_1 decays so fast that we again have $\mathtt{P}(\lambda \in \mathcal{E}_n) \in O(\beta^{-n})$. Then, since $\ell(\mathtt{b}_n) \in O(|n|)$, the quantization rate $\mathcal{R}(\mathtt{q}_\beta) = \sum_{n\in\mathbb{Z}\setminus\{0\}} \ell(\mathtt{b}_n)\mathtt{P}(\lambda \in \mathcal{E}_n) = O(\sum_{n\in\mathbb{Z}\setminus\{0\}} |n|\beta^{-|n|}) = O(1)$ is finite for any $\beta > 1$.

Finally, as $\beta \to \infty$, $\mathtt{P}(\lambda \in \mathcal{E}_{-1}) = \mathtt{P}(\lambda \in [\frac{1}{\beta}, \beta)) \to 1$, and thus, the (finite) rate contributions of all the other quantization cells vanishes. Also, since the codeword $\mathtt{b}_{-1}$ for $\mathcal{E}_{-1}$ has length 1, we have, as $\beta \to \infty$, $\mathtt{R}(\mathtt{q}_\beta) \to \mathtt{P}(\lambda \in \mathcal{E}_{-1})\ell(\mathtt{b}_{-1}) \to 1$. Hence, the role of the deadzone cell $\mathcal{E}_{-1}$ is to make sure that $\mathtt{q}_\beta$ supports feedback rates that are arbitrarily close to 1 bit while still achieving a positive delay-limited capacity with bounded power.

For $m > 1$, the quantizer $\mathtt{q}_\beta$ is roughly the m-fold product of its special case where $m = 1$. The log-uniform intervals for $m = 1$ become log-uniform hypercubes for $m > 1$. Similarly, for a general m, the log-uniform structure provides a capacity of $c(\sigma_\star)$ with finite average power and finite feedback rate. Formal calculations lead to the following result.

Theorem 1. *For every $\beta > 1$ and $P \geq 0$, we have $c(\mathtt{q}_\beta) \geq c(\sigma_\star) = c(P)$, $\mathcal{P}(\mathtt{q}_\beta) \leq \beta^2\mathcal{P}(\sigma_\star) = \beta^2 P$, and $\mathcal{R}(\mathtt{q}_\beta) \leq 1 + \frac{k_1}{\beta-1}$, where k_1 is a constant that depends only on r.*

Proof. See Appendix A. □

We recall that the full-CSIT delay-limited capacity for a given power constraint P is $c(\sigma_\star) = c(P)$. Theorem 1 shows that for any given $P \geq 0$ and $\beta > 1$, we can achieve a delay-limited capacity of $c(P)$ with an average power of $\beta^2 P$ and a feedback rate of $1 + \frac{k_1}{\beta-1}$. Therefore, the quantizer design parameter β controls the tradeoff between the average transmission power and the feedback rate for a fixed target delay-limited capacity. In this context, a large β translates to a large average transmission power but a low feedback rate, while a small β means a low average transmission power but a high feedback rate.

We now investigate the tradeoff between the average feedback rate and the delay-limited capacity for a fixed transmission power. In fact, for any given $R > 1$, applying Theorem 1 for the special case $\beta = 1 + \frac{k_1}{R-1}$, we observe that the capacity with a feedback rate of $R > 1$ is at least $c(P)$ with a power consumption of $(1 + \frac{k_1}{R-1})^2 P$. Equivalently, the capacity $c_R(P)$

with a feedback rate of $R > 1$ and a power constraint of P satisfies

$$c_R(P) \geq c\left(\frac{P}{(1+\frac{k_1}{R-1})^2}\right). \tag{9}$$

Hence, at any feedback rate greater than 1 bit per channel state, and at any power constraint, there is in fact a positive capacity that is achievable by a VLQ. Moreover, by allowing a sufficiently large feedback rate, the capacity with partial-CSIT can be made arbitrarily close to the capacity with full-CSIT.

It is also insightful to compare (9) with the existing results on the capacity of MIMO broadcast channels with limited feedback [38], [39]. It is well-known that in MIMO broadcast channels, to achieve the multiplexing gain of a full-CSIT system, the per-receiver feedback rates should increase logarithmically with the transmission power [38]. The necessity for such high feedback rates is due to the interference between the data streams that are intended for different receivers. On the other hand, (9) shows that in the interference-free point-to-point MIMO channels with the delay-limited capacity performance measure, one can achieve the full-CSIT multiplexing gain with any feedback rate greater than 1 bit per channel state. Hence, in contrast to MIMO broadcast channels, the rate loss due to quantization is bounded uniformly, and moreover, it decays to zero as the feedback rate grows to infinity.

A. Extensions to Multi-User Networks

In this subsection, we extend the above results to multi-user MIMO multicast networks. Suppose that we wish to achieve an outage-free multicast of the symbol $\mathbf{s} \sim \mathcal{CN}(\frac{\sigma}{t}\mathbf{I}_t)$ from the t-antenna transmitter to K receivers with r antennas each. Let $\mathbf{H}_k \in \mathbb{C}^{r\times t}$ denote the channel state from the transmitter to Receiver k, and $\boldsymbol{\lambda}_k = [\lambda_{k1}, \ldots, \lambda_{km}]$ denote the unordered eigenvalues of $\mathbf{H}_k^\dagger \mathbf{H}_k$. We define the delay-limited capacity in this case as the supremum of all data rates that can be reliably decoded at every receiver with zero outage. Using the same arguments as in the single-user case, the corresponding full-CSIT mapping is $\sigma_\star(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K) \triangleq \min\{\sigma : \min_k c(\sigma, \boldsymbol{\lambda}_k) = c'\}$, where c' satisfies $\mathtt{E}[\sigma_\star(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K)] = P$. The delay-limited capacity of the K-user MIMO multicast channel thus equals c'.

For limited-feedback purposes, we use the same encoder as the quantizer $\mathtt{q}_\beta$ at the receivers. In other words, for every k, Receiver k feeds back $\mathtt{b}_{i_k}$ whenever $\boldsymbol{\lambda}_k \in \mathcal{E}_{i_k}$ for some $i_k \in \mathcal{K}$. Using the feedback bits from Receiver k, the transmitter can determine (in the same manner as in the single-user MIMO system) the minimum amount of power to avoid outage at rate c' at Receiver k up to β^2-multiplicative accuracy. Hence, using the feedback bits from all the receivers, the transmitter can determine the minimum amount of power to avoid outage at all the receivers up to β^2-multiplicative accuracy. Therefore, by (9), the capacity of a MIMO multicast system with R bits of feedback per receiver at power level P is at least the capacity of a MIMO multicast system with full-CSIT at power level $P/(1+\frac{k_1}{R-1})^2$.

IV. VLQs for General Precoding Matrices

We now consider the general scenario where the input symbol $\mathbf{s}$ of the MIMO channel is chosen as $\mathbf{s} \sim \mathcal{CN}(\mathbf{Q})$ for some arbitrary covariance matrix $\mathbf{Q} \in \mathbb{C}^{t\times t}$. Equivalently, we can choose $\mathbf{s} \sim \mathbf{X}\mathcal{CN}(\mathbf{I}_t)$, where $\mathbf{X} \in \mathbb{C}^{t\times t}$ satisfies $\mathbf{Q} = \mathbf{X}\mathbf{X}^\dagger$. We call $\mathbf{X}$ a precoding matrix corresponding to $\mathbf{Q}$. For a fixed $\mathbf{H}$, the channel capacity with a precoding matrix $\mathbf{X}$ is

$$C(\mathbf{X},\mathbf{H}) \triangleq \log\det(\mathbf{I}_r + \mathbf{H}\mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger) \text{ nats/sec/Hz}. \tag{10}$$

Again, we assume that the channel codeword spans only one fading block.

Consider a mapping $\mathbf{X} : \mathbb{C}^{r\times t} \to \mathbb{C}^{t\times t}$ that assigns the precoding matrix $\mathbf{X}(\mathbf{H})$ to channel state $\mathbf{H}$. Then, we define the capacity with $\mathbf{X}$ as

$$C(\mathbf{X}) \triangleq \sup\{c \geq 0 : \mathtt{P}(C(\mathbf{X}(\mathbf{H}),\mathbf{H}) < c) = 0\}. \tag{11}$$

Similar to the case of scaled-identity precoding matrices, a "full-CSIT mapping"

$$\mathbf{X}_\star \triangleq \arg\max_{\mathbf{X}}\{C(\mathbf{X}) : \mathtt{E}[\|\mathbf{X}\|^2] \leq P\} \tag{12}$$

exists [1] and provides the maximum achiveable capacity $C(P) \triangleq C(\mathbf{X}_\star)$ given power constraint P. Similarly, we have $\mathbf{X}_\star(\mathbf{H}) = \arg\min_{\mathbf{X}} \{\|\mathbf{X}\| : C(\mathbf{X},\mathbf{H}) = d\}$, where d is chosen to satisfy $\mathtt{E}[\|\mathbf{X}_\star\|^2] = P$.

For a more explicit form of $\mathbf{X}_\star$, consider the decomposition $\mathbf{H}^\dagger\mathbf{H} = \mathbf{U}\mathrm{diag}([\lambda_1 \ \cdots \ \lambda_m \ 0 \ \cdots \ 0])\mathbf{U}^\dagger$, where $\mathbf{U} \in \mathbb{C}^{t\times t}$ is a unitary matrix. Then, the optimal precoding matrix is derived from the water-filling solution [2, Lemma 1]

$$\mathbf{X}_\star = \mathbf{U}\mathrm{diag}\left(\left[\sqrt{(\mu_\star - \tfrac{1}{\lambda_1})^+} \ \cdots \ \sqrt{(\mu_\star - \tfrac{1}{\lambda_m})^+} \ 0 \ \cdots \ 0\right]\right), \tag{13}$$

where

$$\mu_\star \triangleq \left(\frac{e^d}{\prod_{i\in\mathcal{I}_\star}\lambda_i}\right)^{1/|\mathcal{I}_\star|} \tag{14}$$

is usually referred to as the water level, and $\mathcal{I}_\star \subset \{1,\dots,m\}$ is the unique index set that satisfies $\mu_\star \geq \frac{1}{\lambda_i}$ if and only if $i \in \mathcal{I}_\star$. Note that we have omitted (and from now on will omit) the $\mathbf{H}$-dependencies of most quantities (such as $\mathbf{X}_\star$ and $\mu_\star$) for brevity. To gain intuition on how (13) and (14) are obtained (see [2, Lemma 1] for a detailed derivation), first note that for a given channel state, the precoding matrix $\mathbf{X}_\star$ should achieve a mutual information of d with the minimum possible (instantaneous) power. Without loss of generality, we can assume that $\mathbf{X}_\star$ also results in the maximum mutual information among all other precoding matrices with the same power. It then follows from [12] that $\mathbf{X}_\star$ should be of the form (13) for some $\mu_\star > 0$. Using the fact that $\mathbf{X}^\star$ should achieve a mutual information of d, i.e. substituting $\mathbf{X}^\star$ to the equality $C(\mathbf{X},\mathbf{H}) = d$, we have $\prod_{i=1}^m (1+\lambda_i(\mu_\star - \frac{1}{\lambda_i})^+) = e^d$. Solving for $\mu_\star$, we obtain (14). The constant d (and thus also $\mu_\star$) should be chosen according to the long-term power constraint of the transmitter.

For a SIMO system, the most general precoding matrices are scaled-identity matrices. We obtain the mapping $\mathbf{X}_\star = \sqrt{(r-1)P/\lambda_1}$, which is (not surprisingly) equivalent to the mapping $\sigma_\star(\lambda_1) = (r-1)P/\lambda_1$ considered in Section III. In particular, we have $C(P) = c(P)$ when $t = 1$. For a MISO system however, we have $C(P) = c(tP) = \log(1+(t-1)P)$, and therefore, there is a power/beamforming gain of t with a general precoding structure when $r = 1$. For a general r and t, a similar power gain of $\frac{t}{m} \geq 1$ is achievable: We have $C(P) \geq c(\frac{t}{m}P)$ for all $P \geq 0$, and $C(P) = c(\frac{t}{m}P) + o(1)$ as $P \to \infty$ [1].

The goal of this section is therefore to achieve the maximum capacity $C(P)$, which is in general larger than the capacity $c(P)$ with scaled-identity precoding matrices only. Given a feedback rate constraint of R and a power constraint of P, let $C_R(P)$ denote the capacity of the MIMO channel with general quantized precoding matrices. Our main result in this section shows that there is a constant $k > 0$ such that for every sufficiently large R, we have

$$C_R(P) \geq C(P) - \frac{k}{R},\ \forall P \geq 0. \tag{15}$$

In the following, we provide an explicit construction of the VLQs that can achieve the capacities claimed by (15). Our strategy is first to investigate what we call "perturbations" on the optimal precoding matrix $\mathbf{X}_\star$ with full-CSIT. First, we define the notion of "perturbation" for a general matrix and investigate the properties of perturbed matrices.

Definition 1. *Let $\mathbf{X} = [x_{ij}]_{t\times t} \in \mathbb{C}^{t\times t}$ be an arbitrary matrix. We call $\mathbf{Y} = [y_{ij}]_{t\times t} \in \mathbb{C}^{t\times t}$ an*

ϵ-perturbation of $\mathbf{X}$ *if for any* $i, j \in \{1, \dots t\}$*, we have* $|\Re x_{ij} - \Re y_{ij}| \leq \epsilon$ *and* $|\Im x_{ij} - \Im y_{ij}| \leq \epsilon$*.*

Now, given $x \in \mathbb{R}$ and $\xi > 0$, let $\mathtt{Q}(x;\xi) = \mathrm{sign}(x)\xi \lfloor |x|/\xi \rfloor$. We have $|x| \leq |\mathtt{Q}(x;\xi)|$. In fact, $|x| - |\mathtt{Q}(x;\xi)| \leq \xi$, and therefore, $\mathtt{Q}(x;\xi)$ is a ξ-perturbation of x. We extend the definition of $\mathtt{Q}(x;\xi)$ to an arbitrary complex matrix $\mathbf{A} = [a_{ij}]_{M\times N}$ via defining $\mathtt{Q}(\mathbf{A};\xi) = [\mathtt{Q}(\Re a_{ij};\xi) + \sqrt{-1}\mathtt{Q}(\Im a_{ij};\xi)]_{M\times N}$. The following lemma immediately follows from these definitions.

Lemma 1. *For any* $\xi > 0$ *and any matrix* $\mathbf{A}$*, the (quantized) matrix* $\mathtt{Q}(\mathbf{A};\xi)$ *is a* ξ*-perturbation of* $\mathbf{A}$ *with* $\|\mathtt{Q}(\mathbf{A};\xi)\| \leq \|\mathbf{A}\|$*. Moreover, in the domain of all* $M \times N$ *matrices* $\mathbf{A}$ *with* $\|\mathbf{A}\| \leq a$*, the matrix* $\mathtt{Q}(\mathbf{A};\xi)$ *can assume at most* $(1+\frac{2a}{\xi})^{2MN}$ *values. In other words, the cardinality bound* $|\{\mathtt{Q}(\mathbf{A};\xi) : \mathbf{A} \in \mathbb{C}^{M\times N},\ \|\mathbf{A}\| \leq a\}| \leq (1+\frac{2a}{\xi})^{2MN}$ *holds for every* $a \geq 0$*.*

Proof. The first sentence in the statement of the lemma is a straightforward consequence of the definitions. For the cardinality bound, note that if $|x| \leq a$ for some $a > 0$, then $q(x;\xi)$ can assume at most $2\lfloor \frac{a}{\xi} \rfloor + 1 \leq 1 + \frac{2a}{\xi}$ values. On the other hand, $\|\mathbf{A}\| \leq a$ implies that the real and imaginary parts of all the MN components of $\mathbf{A}$ have norms no greater than a. Hence, when $\|\mathbf{A}\| \leq a$, the matrix $\mathtt{Q}(\mathbf{A};\xi)$ can assume at most $(1+\frac{2a}{\xi})^{2MN}$ values. □

In the following, we find an upper bound on the data rate loss due to perturbations of an optimal precoding matrix. For future reference, we first prove a more general result.

Proposition 1. *For a given* $\mathbf{H}$ *with* $\mathbf{H}^\dagger\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$*, let* $\mathbf{X} = \mathbf{U}\sqrt{\mathbf{N}}$*, where* $\mathbf{N} = \mathrm{diag}((\nu - \frac{1}{\lambda_1})^+, \dots, (\nu - \frac{1}{\lambda_m})^+, 0, \dots, 0) \in \mathbb{C}^{t\times t}$ *for some arbitrary* $\nu \in \mathbb{R}$*. Then, for every* $\epsilon\sqrt{\nu}$*-perturbation* $\widehat{\mathbf{X}}$ *of* $\mathbf{X}$*, we have*

$$C(\widehat{\mathbf{X}}, \mathbf{H}) \geq C(\mathbf{X}, \mathbf{H}) - k_2\epsilon \tag{16}$$

for every sufficiently small ϵ *and some constant* k_2*.*

Proof. See Appendix B. □

For our specific problem at hand, we need the following corollary.

Corollary 1. *For a given* $\mathbf{H}$*, let* $\mathbf{X}_\star$ *be as defined in (13)-(14). Then, for every* $\epsilon\sqrt{\mu_\star}$*-perturbation* $\widehat{\mathbf{X}}$ *of* $\mathbf{X}_\star$*, we have* $C(\widehat{\mathbf{X}}, \mathbf{H}) \geq C(P) - k_2\epsilon$ *for some constant* $k_2 > 0$ *and every sufficiently small* ϵ*.*

Proof. By Proposition 1, we have $C(\widehat{\mathbf{X}}, \mathbf{H}) \geq C(\mathbf{X}_\star, \mathbf{H}) - k_2\epsilon = C(P) - k_2\epsilon$ as desired. □

Suppose now that the transmitter knows $\mu_\star$ perfectly. Then, Corollary 1 and Lemma 1 lead to the following simple quantization scheme that utilizes the quantized precoding matrix $\mathbf{X}' \triangleq \mathtt{Q}(\mathbf{X}_\star; \epsilon\sqrt{\mu_\star})$. By (13), we have $\|\mathbf{X}_\star\| \leq t\sqrt{\mu_\star}$, and therefore, by Lemma 1, $\mathbf{X}'$ can assume at most $(1 + \frac{2t}{\epsilon})^{2t^2}$ values. In other words, $\mathbf{X}'$ can be represented by a rate-$R(\epsilon)$ FLQ, where $R(\epsilon) \triangleq 2t^2 \log_2(1 + \frac{2t}{\epsilon})$. Hence, the receiver, knowing $\mu_\star$, quantizes the optimal matrix $\mathbf{X}_\star$ to $\mathbf{X}'$ and sends the corresponding $R(\epsilon)$ feedback bits to the transmitter. Transmitter, knowing $\mu_\star$, can recover $\mathbf{X}'$. By Lemma 1, $\mathbf{X}'$ is an $\epsilon\sqrt{\mu_\star}$-perturbation of $\mathbf{X}_\star$, and therefore, it provides a capacity of at least $C(P) - k_2\epsilon$ by Corollary 1. Also, since $\|\mathbf{X}'\| \leq \|\mathbf{X}_\star\|$ by Lemma 1, and $\mathtt{E}[\|\mathbf{X}_\star\|^2] = P$, the average power consumption with $\mathbf{X}'$ is at most P. Therefore, we can achieve the capacity $C(P) - k_2\epsilon$ with a power consumption of P and a (fixed-length) feedback rate of $R(\epsilon)$.

Obviously, the transmitter cannot know $\mu_\star$ perfectly. Here, we observe that we can at least tolerate a "multiplicative uncertainty" regarding the knowledge of $\mu_\star$ at the transmitter. To be more precise, for $\mu > 0$ and $\zeta \geq 1$, let $\mathcal{H}_{\mu,\zeta} \triangleq \{\mathbf{H} : \frac{\mu}{\zeta} \leq \mu_\star \leq \mu\}$. Then, we have the following result.

Proposition 2. *For every $\mu > 0$ and $\zeta \geq 1$, the following holds.*

(i) For every $\mathbf{H} \in \mathcal{H}_{\mu,\zeta}$, the matrix $\mathtt{Q}\Big(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}}\Big)$ is an $\epsilon\sqrt{\mu_\star}$-perturbation of $\mathbf{X}_\star$.

(ii) For every $\mathbf{H} \in \mathcal{H}_{\mu,\zeta}$, we have $\|\mathtt{Q}\Big(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}}\Big)\| \leq \|\mathbf{X}_\star\|$.

(iii) The bound $\Big|\Big\{\mathtt{Q}\Big(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}}\Big) : \mathbf{H} \in \mathcal{H}_{\mu,\beta}\Big\}\Big| \leq 2^{R(\epsilon/\sqrt{\zeta})}$ holds.

Proof. For (i), note that $\mathtt{Q}\Big(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}}\Big)$ is an $\epsilon\sqrt{\frac{\mu}{\zeta}}$-perturbation of $\mathbf{X}_\star$ by definition. It is also an $\epsilon\sqrt{\mu_\star}$-perturbation of $\mathbf{X}_\star$ as $\mu_\star \geq \frac{\mu}{\zeta}$ whenever $\mathbf{H} \in \mathcal{H}_{\mu,\zeta}$. (ii) holds by the definition of $\mathtt{Q}(\cdot)$. For (iii), first note that the real and imaginary parts of all the t^2 components of $\mathbf{X}_\star$ have norm no greater than $t\sqrt{\mu_\star} \leq t\sqrt{\mu}$. It follows that $q\Big(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}}\Big)$ can assume at most

$$\left[2\left\lfloor\frac{t\sqrt{\mu}}{\epsilon\sqrt{\frac{\mu}{\zeta}}}\right\rfloor + 1\right]^{2t^2} = \left[2\left\lfloor\frac{t\sqrt{\zeta}}{\epsilon}\right\rfloor + 1\right]^{2t^2} \leq \left[2\frac{t\sqrt{\zeta}}{\epsilon} + 1\right]^{2t^2} = 2^{R(\epsilon/\sqrt{\zeta})} \tag{17}$$

values. □

Hence, if the transmitter only knows that $\frac{\mu}{\zeta} \leq \mu_\star \leq \mu$ for some $\mu > 0$ and $\zeta \geq 1$ (instead of a perfect knowledge of $\mu_\star$), we can utilize the quantized precoding matrix $\mathtt{Q}(\mathbf{X}_\star; \epsilon\sqrt{\frac{\mu}{\zeta}})$ to achieve a capacity of $C(P) - k_2\epsilon$ with a power consumption of P. The only difference is that the required feedback rate will be $R(\epsilon/\sqrt{\zeta})$ instead of $R(\epsilon)$. The remaining issue is therefore to quantize $\mu_\star$ to (say) $\widehat{\mu}$ in such a way that the inequalities $\frac{\widehat{\mu}}{\zeta} \leq \mu_\star \leq \widehat{\mu}$ always hold. This can be accomplished using the log-uniform quantizer structure discussed in Section III.

For $\boldsymbol{\lambda}' = [\lambda_1' \cdots \lambda_m']$, let $\mu_\star(\boldsymbol{\lambda}')$ be defined as in (14) with respect to $\lambda_1', \ldots, \lambda_m'$. Let $\overline{\mu}_{\{-1\}^m} = \mu_\star([\frac{1}{\beta} \cdots \frac{1}{\beta}])$, and for $[i_1 \cdots i_m] \in \mathcal{K} - \{0\}$, let $\overline{\mu}_{[i_1 \cdots i_m]} \triangleq \mu_\star([\beta^{i_1} \cdots \beta^{i_m}])$. Let $\widehat{\mu} = \overline{\mu}_\kappa$ and $\widehat{\mathtt{b}} = \mathtt{b}_\kappa$ whenever $\boldsymbol{\lambda} \in \mathcal{E}_\kappa$. We have the following.

Proposition 3. *For every $\beta \geq 1$, we have $\frac{\widehat{\mu}}{\beta^2} \leq \mu_\star \leq \widehat{\mu}$.*

Proof. See Appendix C. □

We can now consider a quantization scheme that operates in the following manner. For a given $\mathbf{H}$, the receiver first calculates $\mu_\star$ and quantizes $\mu_\star$ to $\widehat{\mu}$. The feedback bits for this variable-length stage is given by the binary codeword $\widehat{\mathtt{b}}$. Then, having calculated $\widehat{\mu}$, the receiver quantizes the optimal precoding matrix $\mathbf{X}_\star$ to $\mathtt{Q}(\mathbf{X}_\star; \epsilon\sqrt{\widehat{\mu}/\beta^2})$. This fixed-length stage requires $R(\frac{\epsilon}{\beta})$ feedback bits. The receiver first sends the feedback bits for the fixed-length stage followed by the feedback bits $\widehat{\mathtt{b}}$ for the variable-length stage (so that we have a prefix-free code). The average feedback rate is no more than $R(\frac{\epsilon}{\beta}) + 1 + \frac{k_1}{\beta - 1}$ by Theorem 1. The transmitter, on the other hand, first recovers $\widehat{\mu}$, and then reconstructs and uses $\mathtt{Q}(\mathbf{X}_\star; \epsilon\sqrt{\widehat{\mu}/\beta^2})$. According to Propositions 2, 3, and 4, this provides a capacity of at least $C(P) - k_2\epsilon$ with a power consumption of no more than P. In particular, choosing $\beta = 1 + \frac{k_3}{R}$ and $\epsilon = \frac{k_3}{R}$ for some suitable constant $k_3 > 0$ proves (15). We now formally present the main result of this section as a theorem.

Theorem 2. *There is a constant $k > 0$ such that for every sufficiently large R, we have $C_R(P) \geq C(P) - \frac{k}{R}$ for every $P \geq 0$.*

V. Extensions to B-Block Fading Channels

The two previous sections were concerned with the limited-feedback delay-limited capacities of "1-block fading channels," i.e. fading channels for which the channel codeword is assumed to span only one fading block. In this section, we extend our results to B-block fading channels. As discussed in [1], in a B-block fading channel with channel matrices $\mathbf{H}_1, \ldots, \mathbf{H}_B$, one considers B precoding matrices, $\mathbf{X}_1, \ldots, \mathbf{X}_B$ that achieve the rate

$$\frac{1}{B} \sum_{b=1}^{B} \log\det(\mathbf{I}_r + \mathbf{H}_b \mathbf{X}_b \mathbf{X}_b^\dagger \mathbf{H}_b^\dagger) \tag{18}$$

with power $\frac{1}{B}\sum_{b=1}^{B}\|\mathbf{X}_b\|^2$. Let $\mathbf{H}_b = \mathbf{U}_b\mathrm{diag}([\lambda_{b1} \ \cdots \ \lambda_{bm} \ \ 0 \ \cdots \ 0])\mathbf{U}_b^\dagger$ denote the spectral decomposition of the channel matrix with index b. Using the same arguments as in the case $B = 1$, the optimal collection of precoding matrices $\mathbf{X}_{1\star}, \ldots, \mathbf{X}_{B\star}$ that maximize the capacity are then given by

$$\mathbf{X}_{b\star} = \mathbf{U}_b\mathrm{diag}\left(\left[\sqrt{(\mu_\star - \tfrac{1}{\lambda_{b1}})^+} \ \cdots \ \sqrt{(\mu_\star - \tfrac{1}{\lambda_{bm}})^+} \ 0 \ \cdots \ 0\right]\right),\ b = 1, \ldots, B, \tag{19}$$

with[1]

$$\mu_\star = \left(\frac{e^{dB}}{\prod_{[b\ i]\in\mathcal{I}_\star}\lambda_{bi}}\right)^{1/|\mathcal{I}_\star|}. \tag{20}$$

Here, the constant d is chosen to satisfy the long-term power constraint of the transmitter, and $\mathcal{I}_\star \subset \{1, \ldots, B\} \times \{1, \ldots, m\}$ is the unique index set that satisfies $\mu_\star \geq \frac{1}{\lambda_{bi}}$ if and only if $[b\ i] \in \mathcal{I}_\star$. The derivations of (19) and (20) follow the exact same steps as the derivations of (13) and (14) for the single-block case. The corresponding B-block delay-limited capacity with full-CSIT is

$$C(P;B) \triangleq \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{m}\mathtt{E}\left[(\log(\mu_\star\lambda_{bi}))^+\right]. \tag{21}$$

Now, let $C_R(P;B)$ denote the capacity with rate R quantized CSI per block (so that we use a total of RB feedback bits on average to quantize the collection $\mathbf{H}_1, \ldots, \mathbf{H}_B$.). We wish to generalize Theorem 2 by determining how $C_R(P;B)$ relates to $C(P;B)$ for a general B.

We can observe that the expressions for the optimal precoding matrices for a B-block channel are in the exact same form as the optimal precoding matrix for the single-block case discussed in the previous section. The only difference is that now the power control parameter $\mu_\star$ depends on all the eigenvalues of all the B channels, but still, the nature of the dependence is the same. Hence, the arguments we have used to prove Theorem 2 applies verbatim. We have the following theorem.

Theorem 3. *There is a constant $k > 0$ such that for every $B \geq 1$ and every sufficiently large R, we have $C_R(P;B) \geq C(P;B) - \frac{k}{R}$ for every $P \geq 0$.*

Proof. We can first assume that the transmitter knows $\mu_\star$ perfectly, in which case the sequence of quantized precoding matrices $\mathtt{Q}(\mathbf{X}_{b\star}; \epsilon\sqrt{\mu_\star})$, $b = 1, \ldots, B$ provides a capacity

[1] In order to avoid cumbersome expressions, we use the same notation as in Section IV for variables such as $\mu_\star$. The B-dependence of these variables should be clear from the context.

of at least $C(P;B) - k_2\epsilon$ by Lemma 1 and Proposition 1. The quantization of $\mu_\star$ can be accomplished using the same variable-length strategy described in Section III, with the only difference being that the quantizer of the eigenvalues should now be Bm dimensional instead of m dimensional. These constructions result in a feedback rate amplification of at most B (compared to the case $B = 1$), and thus the delay-limited capacity $C_R(P;B)$ with a feedback rate of R bits per block satisfies $C_R(P;B) \geq C(P;B) - \frac{k}{R}$ for every B, where k is the same constant as stated in Theorem 2. □

VI. Applications to Ergodic Channels

The quantizer constructions and the technical tools we have developed so far can be applied to a variety of limited feedback problems that are concerned with the capacity of MIMO systems. As side results, we discuss here applications to the limited-feedback ergodic capacities of MIMO channels with both long-term and short-term power constraints. Note that the ergodic channel can be considered as a B-block fading channel with $B \to \infty$.

A. *Ergodic Capacity with a Long-Term Power Constraint*

Consider the now-ergodic MIMO channel $\mathbf{H}$ with spectral decomposition $\mathbf{H}^\dagger\mathbf{H} = \mathbf{U}\mathrm{diag}([\lambda_1 \cdots \lambda_m \; 0 \; \cdots \; 0])\mathbf{U}^\dagger$. The ergodic capacity with full-CSIT (c.f. (21))

$$C(P;\infty) \triangleq \sum_{i=1}^{m} \mathtt{E}\left[(\log(\mu_\star \lambda_i))^+\right] \tag{22}$$

is achieved by transmitting via the precoding matrix

$$\mathbf{X}_\star = \mathbf{U}\mathrm{diag}\left(\left[\sqrt{(\mu_\star - \tfrac{1}{\lambda_1})^+} \;\; \cdots \;\; \sqrt{(\mu_\star - \tfrac{1}{\lambda_m})^+} \; 0 \;\; \cdots \;\; 0\right]\right), \tag{23}$$

where the water level $\mu_\star$ is now the solution to the equation

$$\sum_{i=1}^{m} \mathtt{E}\left[(\mu_\star - \tfrac{1}{\lambda_i})^+\right] = P. \tag{24}$$

One important observation is that, as shown by (24), the water level $\mu_\star$ is independent of the channel realization $\mathbf{H}$ in the ergodic scenario. This is in contrast to the finite-block cases we have discussed in the previous sections where $\mu_\star$ has exhibited dependency on the specific channel realization(s). In these finite-block cases, what made variable-length feedback necessary was precisely this channel state dependency and the associated unboundedness of the water level $\mu_\star$. In the ergodic scenario, since $\mu_\star$ becomes a fixed constant (that could be made a-priori available to both the transmitter and the receiver), it follows from Proposition 1 that one can approach the full-CSIT capacity using fixed-length feedback only. Formally, let $C_R(P;\infty)$ denote the ergodic capacity with a long-term power constraint P and a fixed-length feedback rate of R bits per channel state. Then, we have the following theorem.

Theorem 4. *There is a constant $k_4 > 0$ such that for every sufficiently large R, we have $C_R(P;\infty) \geq C(P;\infty) - k_4 2^{-\frac{R}{2t^2}}$ for every $P \geq 0$.*

Proof. By Proposition 1, for every given $\mathbf{H}$, any $\epsilon\sqrt{\mu_\star}$-perturbation of the optimal precoding matrix $\mathbf{X}_\star$ for $\mathbf{H}$ provides a capacity of at least $C(\mathbf{X}_\star, \mathbf{H}) - k_2\epsilon$. As discussed after Corollary 1 in Section IV, finding such a perturbed matrix is always possible as long as we have a codebook of cardinality $(1 + \frac{2t}{\epsilon})^{2t^2} \in O(\epsilon^{-2t^2})$. This means that we can achieve an ergodic capacity of $\mathtt{E}[C(\mathbf{X}_\star, \mathbf{H}) - k_2\epsilon] = C(P;\infty) - O(\epsilon)$ with a codebook of cardinality $O(\epsilon^{-2t^2})$. Letting $\epsilon = 2^{-\frac{R}{2t^2}}$ concludes the proof. □

Hence, variable-length feedback is not necessary in the case of ergodic channels, and one can approach the full-CSIT ergodic capacity using fixed-length feedback only. Also, recall that in the non-ergodic case, the difference between the quantized-CSIT and the full-CSIT capacity decays at least inverse linearly with the feedback rate. In the ergodic case, the difference decays at least exponentially with the feedback rate.

B. Ergodic Capacity with a Short-Term Power Constraint

Finally, we consider the limited-feedback ergodic capacity of the MIMO channel with a short-term power constraint P. This problem was previously studied in [27] via random vector quantizers and it was shown that the capacity loss due to fixed-length quantization decays at least exponentially with the feedback rate. Here, we obtain the same result using the structured non-random quantizers that we have introduced in the previous sections. Hence, we can provide a simple explicit construction of a "good" quantizer that provides an exponentially-decaying quantization loss.

For a MIMO channel with a short-term power constraint P, the ergodic capacity

$$\widetilde{C}(P) \triangleq \sum_{i=1}^{m} \mathtt{E}\left[(\log(\widetilde{\mu}\lambda_i))^+\right] \tag{25}$$

with full-CSIT is achieved by the precoding matrix

$$\widetilde{\mathbf{X}} \triangleq \mathbf{U}\mathrm{diag}\left(\left[\sqrt{(\widetilde{\mu} - \tfrac{1}{\lambda_1})^+} \ \cdots \ \sqrt{(\widetilde{\mu} - \tfrac{1}{\lambda_m})^+} \ 0 \ \cdots \ 0\right]\right), \tag{26}$$

where the water level $\widetilde{\mu}$ satisfies

$$\sum_{i=1}^{m}(\widetilde{\mu}-\tfrac{1}{\lambda_i})^+ = P. \tag{27}$$

Unlike the long-term ergodic scenario discussed in Section VI-A, the optimal water level now depends on $\mathbf{H}$. On the other hand, (27) implies $\widetilde{\mu} \geq \frac{P}{t}$, and thus the water level is bounded from below for every $\mathbf{H}$. Since we also have $\|\widetilde{\mathbf{X}}\| = P$ for every $\mathbf{H}$, a fixed-length feedback strategy as suggested by Proposition 1 again becomes feasible. Formally, let $\widetilde{C}_R(P)$ denote the ergodic capacity with a short-term power constraint P and a fixed-length feedback rate of R bits per channel state. We have the following theorem.

Theorem 5. *There is a constant $k_5 > 0$ such that for every sufficiently large R, we have $\widetilde{C}_R(P) \geq \widetilde{C}(P) - k_5 2^{-\frac{R}{2t^2}}$ for every $P \geq 0$.*

Proof. Consider the quantized precoding matrix $\mathtt{Q}(\widetilde{\mathbf{X}}; \epsilon\sqrt{\frac{P}{t}})$, which, by definition, is an $\epsilon\sqrt{\frac{P}{t}}$-perturbation of $\widetilde{\mathbf{X}}$. Since $\widetilde{\mu} \geq \frac{P}{t}$, it is also an $\epsilon\sqrt{\widetilde{\mu}}$-perturbation of $\widetilde{\mathbf{X}}$ and thus achieves a capacity of at least $C(\widetilde{\mathbf{X}}, \mathbf{H}) - O(\epsilon)$ by Proposition 1. Moreover, since $\|\widetilde{\mathbf{X}}\| = \sqrt{P}$, the matrix $\mathtt{Q}(\widetilde{\mathbf{X}}; \epsilon\sqrt{\frac{P}{t}})$ can assume at most $(1+\frac{2\sqrt{t}}{\epsilon})^{2t^2} = O(\epsilon^{-2t^2})$ values by Lemma 1. Therefore, averaging out over all the channels, we can achieve an ergodic capacity of $\widetilde{C}(P) - O(\epsilon)$ using $O(\epsilon^{-2t^2})$ quantized precoding matrices. Letting $\epsilon = 2^{-\frac{R}{2t^2}}$ concludes the proof. □

VII. Conclusions

We have studied the limited feedback delay-limited capacities of a fixed data rate slow-fading MIMO channel with a long-term power constraint P at the transmitter. In this context, the standard partial CSIT model where the transmitter has a fixed finite bits of quantized CSI feedback for each channel state results in zero delay-limited capacity. We have shown that by using a variable-length feedback scheme that utilizes different number of feedback bits for different channel states, a non-zero delay-limited capacity can be achieved if the feedback rate is greater than 1 bit per channel state. We have also shown that the delay-limited capacity loss due to finite-rate feedback decays at least inverse linearly with respect to the feedback rate. We have extended our results to B-block fading channels, and to ergodic MIMO channels where one lets $B \to \infty$. For the latter scenario, we have shown that the ergodic capacity loss due to quantization decays at least exponentially with the feedback rate and fixed-length feedback is sufficient to achieve this performance.

We have thus presented a general unifying treatment of the limited-feedback capacities of fixed data rate B-block MIMO fading channels from $B = 1$ all the way up to $B = \infty$. For finite B, we have shown that variable-length feedback with power control is necessary and sufficient to achieve a positive capacity and to approach the capacity with perfect CSIT. For $B = \infty$, fixed-length feedback is sufficient to approach the capacity with perfect CSIT.

One complication in implementing variable-length feedback schemes is the fact that most current wireless systems are designed with fixed-length feedback in mind. Nevertheless, our results demonstrate that variable-length feedback can provide fundamental performance gains over fixed-length feedback, especially for delay-limited applications and power control. Hence, there is a need for designing new or modifying existing wireless communication protocols so as to enable variable-length resource allocation in the feedback link. Some initial work in this context for 802.11 networks can be found in [42].

Appendix A
Proof of Theorem 1

We first prove that $c(\mathsf{q}_\beta) \geq c(\sigma_\star)$ and $\mathcal{P}(\mathsf{q}_\beta) \leq \beta^2 \mathcal{P}(\sigma_\star)$. We need the following lemma.

Lemma 2. *Let $\boldsymbol{\lambda}' = [\lambda'_1 \cdots \lambda'_m] \in \mathbb{R}^m_{>0}$ and $\beta' \geq 1$. Then, for any $\boldsymbol{\lambda}$ with $\lambda'_i \leq \lambda_i \leq \beta' \lambda'_i$ for every $i \in \{1, \ldots, m\}$, we have $\sigma_\star(\boldsymbol{\lambda}) \leq \sigma_\star(\boldsymbol{\lambda}') \leq \beta' \sigma_\star(\boldsymbol{\lambda}')$.*

Proof. Let $\sigma = \sigma_\star(\boldsymbol{\lambda})$ and $\sigma' = \sigma_\star(\boldsymbol{\lambda}')$. We have $c(\sigma, \boldsymbol{\lambda}) = c(\sigma', \boldsymbol{\lambda}') = c$ for some $c > 0$. Note that in general, $c(\sigma'', \boldsymbol{\lambda}'')$ is an increasing function of σ'' and the components of $\boldsymbol{\lambda}''$. In particular, since $\lambda_i \geq \lambda'_i$ for all i, we obtain $c(\sigma', \boldsymbol{\lambda}) \geq c(\sigma', \boldsymbol{\lambda}')$. But since we also have $c(\sigma', \boldsymbol{\lambda}') = c(\sigma, \boldsymbol{\lambda})$, we obtain $c(\sigma', \boldsymbol{\lambda}) \geq c(\sigma, \boldsymbol{\lambda})$. This implies $\sigma' \geq \sigma$ and proves the upper bound on $\sigma_\star(\boldsymbol{\lambda})$. The proof for the lower bound on $\sigma_\star(\boldsymbol{\lambda})$ is similar and is thus omitted. □

This leads to the following bounds on the power consumption of q_β.

Proposition 4. *For every $m \geq 1$, $\beta > 1$, and $\boldsymbol{\lambda} \in \mathbb{R}_{>0}$, we have $\sigma_\star(\boldsymbol{\lambda}) \leq \mathsf{q}_\beta(\boldsymbol{\lambda}) \leq \beta^2 \sigma_\star(\boldsymbol{\lambda})$.*

Proof. Letting $\boldsymbol{\lambda}' = [\frac{1}{\beta} \cdots \frac{1}{\beta}]$ and $\beta' = \beta^2$ in Lemma 2, we obtain $\sigma_\star(\boldsymbol{\lambda}) \leq \mathsf{q}_\beta(\boldsymbol{\lambda}) = \sigma_\star([\frac{1}{\beta} \cdots \frac{1}{\beta}]) \leq \beta^2 \sigma_\star(\boldsymbol{\lambda})$, which proves the claim when $\boldsymbol{\lambda} \in \mathcal{E}_{\{-1\}^m}$. Similarly, letting $\boldsymbol{\lambda}' = [\beta^{i_1} \cdots \beta^{i_m}]$ for $[i_1 \cdots i_m] \in \mathcal{K} \backslash \{0\}$ with $\beta' = \beta$, we obtain $\sigma_\star(\boldsymbol{\lambda}) \leq \mathsf{q}_\beta(\boldsymbol{\lambda}) = \sigma_\star([\frac{1}{\beta} \cdots \frac{1}{\beta}]) \leq \beta \sigma_\star(\boldsymbol{\lambda}) \leq \beta^2 \sigma_\star(\boldsymbol{\lambda})$ for $\boldsymbol{\lambda} \in \mathcal{E}_{[i_1 \cdots i_m]}$, and this concludes the proof. □

In particular, the lower bound on $\mathtt{q}_\beta(\boldsymbol{\lambda})$ in Proposition 4 implies $c(\mathtt{q}_\beta) \geq c(\sigma_\star)$, and the upper bound on $\mathtt{q}_\beta(\boldsymbol{\lambda})$ in Proposition 4 implies $\mathcal{P}(\mathtt{q}_\beta) \leq \beta^2 \mathcal{P}(\sigma_\star)$. What is left is now to prove the upper bound on $\mathcal{R}(\mathtt{q}_\beta)$.

Let $\boldsymbol{\Lambda}$ represent the random vector corresponding to the eigenvalues $\boldsymbol{\lambda}$. We have

$$f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) = K_{m,\overline{m}}^{-1} \exp\left(-\sum_{i=1}^m \lambda_i\right) \prod_{i=1}^m \lambda_i^{\overline{m}-m} \prod_{\substack{i<j \\ i,j\in\{1,\ldots,m\}}} (\lambda_i - \lambda_j)^2, \tag{28}$$

where $K_{m,\overline{m}}$ is a normalizing constant. Note that the second product is equal to 1 when $m = 1$. For $m \geq 2$, we have $(\lambda_i - \lambda_j)^2 \leq 2(\lambda_i^2 + \lambda_j^2) \leq 2\sum_{i=1}^m \lambda_i^2$. This implies $\prod_{i<j}(\lambda_i - \lambda_j)^2 \leq_c 2^{m^2}(\sum_i \lambda_i^2)^{m^2} \leq_c \sum_i \lambda_i^{2m^2}$, where "$\leq_c$" means that the inequality holds up to a constant multiplier that depends only on t and r. Hence, for any $m \geq 1$, we have $\prod_{i<j}(\lambda_i - \lambda_j)^2 \leq_c 1 + \sum_i \lambda_i^{2m^2}$, and therefore, the joint PDF of $\lambda_1, \ldots, \lambda_m$ admits an upper bound of the form

$$f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) \leq_c \sum_{i=1}^{m+1} \prod_{j=1}^m e^{-\lambda_j} \lambda_j^{a_{ij}-1} \tag{29}$$

for constants $1 \leq a_{ij} \leq_c 1$. Now, let $Z_{ij} \sim \Gamma(a_{ij}, 1)$, $i = 1, \ldots, m+1$, $j = 1, \ldots, m$ be $m(m+1)$ independent Gamma random variables. Let us also define the random vector $\mathbf{Z}_i = [Z_{i1} \cdots Z_{im}]$. We have

$$f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) \leq_c \sum_{i=1}^{m+1} \prod_{j=1}^m f_{Z_{ij}}(\lambda_j) = \sum_{i=1}^{m+1} f_{\mathbf{Z}_i}(\boldsymbol{\lambda}). \tag{30}$$

For a given $\lambda \in \mathbb{R}_{>0}$, let $\mathfrak{l}(\lambda) = |\mathtt{a}_i|$ whenever $\lambda \in [\beta^i, \beta^{i+1})$ for some $i \in \mathbb{Z}$. Also, for a given $\boldsymbol{\lambda} \in \mathbb{R}_{>0}^m$, let $\mathfrak{L}(\boldsymbol{\lambda}) = |\mathtt{b}_\kappa|$ whenever $\boldsymbol{\lambda} \in \mathcal{E}_\kappa$ for some $\kappa \in \mathcal{K}$. Therefore, when $\boldsymbol{\lambda} \in \mathcal{E}_{\{-1\}^m}$, we have $\mathfrak{L}(\boldsymbol{\lambda}) = |\mathtt{b}_{\{-1\}^m}| = |\mathtt{0}| = 1$. Otherwise, $\boldsymbol{\lambda} \in \mathcal{E}_{[i_1 \cdots i_m]}$ for some $[i_1 \cdots i_m] \in \mathbb{Z}^m \backslash \{-1, 0\}^m$, and we have $\mathfrak{L}(\boldsymbol{\lambda}) = |\mathtt{1}\mathtt{a}_{i_1} \cdots \mathtt{a}_{i_m}| = 1 + \sum_{i=1}^m \mathfrak{l}(\lambda_i)$. We now have

$$\begin{aligned}
\mathcal{R}(\mathtt{q}_\beta) &= \mathtt{E}[\mathfrak{L}(\boldsymbol{\Lambda})] && (31)\\
&= \underbrace{\mathtt{E}[\mathfrak{L}(\boldsymbol{\Lambda})|\boldsymbol{\Lambda} \in \mathcal{E}_{\{-1\}^m}]}_{=1} \underbrace{\mathtt{P}(\boldsymbol{\Lambda} \in \mathcal{E}_{\{-1\}^m})}_{\leq 1} + \mathtt{E}[\mathfrak{L}(\boldsymbol{\Lambda})|\boldsymbol{\Lambda} \in \mathcal{E}_{\{-1\}^m}^c]\mathtt{P}(\boldsymbol{\Lambda} \in \mathcal{E}_{\{-1\}^m}^c) && (32)\\
&= 1 + \int_{\mathcal{E}_{\{-1\}^m}^c} \mathfrak{L}(\boldsymbol{\lambda}) f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) \mathrm{d}\boldsymbol{\lambda} && (33)\\
&\leq_c 1 + \sum_{i=1}^{m+1} \int_{\mathcal{E}_{\{-1\}^m}^c} \mathfrak{L}(\boldsymbol{\lambda}) f_{\mathbf{Z}_i}(\boldsymbol{\lambda}) \mathrm{d}\boldsymbol{\lambda} && (34)\\
&= 1 + \sum_{i=1}^{m+1} \mathtt{E}[\mathfrak{L}(\mathbf{Z}_i)|\mathbf{Z}_i \in \mathcal{E}_{\{-1\}^m}^c]\mathtt{P}(\mathbf{Z}_i \in \mathcal{E}_{\{-1\}^m}^c) && (35)
\end{aligned}$$

Since

$$\mathcal{E}^c_{\{-1\}^m} = \mathbb{R}^m_{>0} - [\tfrac{1}{\beta}, \beta)^m = \bigcup_{j=1}^{m} (\{\mathbf{x} \in \mathbb{R}^m_{>0} : x_j \geq \beta\} \cup \{\mathbf{x} \in \mathbb{R}^m_{>0} : x_j < \tfrac{1}{\beta}\}), \tag{36}$$

we obtain

$$\mathcal{R}(\mathtt{q}_\beta) \leq 1 + \sum_{i=1}^{m+1} \sum_{j=1}^{m} \Big(\mathtt{E}[\mathfrak{L}(\mathbf{Z}_i) | Z_{ij} \geq \beta] \mathtt{P}(Z_{ij} \geq \beta) + \mathtt{E}[\mathfrak{L}(\mathbf{Z}_i) | Z_{ij} < \tfrac{1}{\beta}] \mathtt{P}(Z_{ij} < \tfrac{1}{\beta}) \Big). \tag{37}$$

Let us recall that for any $i \in \{1, \ldots, m\}$, we have $\mathfrak{L}(\mathbf{Z}_i) = 1 + \sum_{k=1}^{m} \mathfrak{l}(Z_{ik})$ whenever $\mathbf{Z}_i \in \mathcal{E}^c_{\{-1\}^m}$. Therefore,

$$\begin{aligned}
\mathcal{R}(\mathtt{q}_\beta) \leq 1 + \sum_{i=1}^{m+1} \sum_{j=1}^{m} \mathtt{E}\left[1 + \sum_{k=1}^{m} \mathfrak{l}(Z_{ik}) \middle| Z_{ij} \geq \beta \right] \mathtt{P}(Z_{ij} \geq \beta) + \\
\sum_{i=1}^{m+1} \sum_{j=1}^{m} \mathtt{E}\left[1 + \sum_{k=1}^{m} \mathfrak{l}(Z_{ik}) \middle| Z_{ij} < \tfrac{1}{\beta} \right] \mathtt{P}(Z_{ij} < \tfrac{1}{\beta})
\end{aligned} \tag{38}$$

$$\begin{aligned}
= 1 + \sum_{i=1}^{m+1} \sum_{j=1}^{m} \left(1 + \mathtt{E}[\mathfrak{l}(Z_{ij}) | Z_{ij} \geq \beta] + \sum_{k \neq j} \mathtt{E}[\mathfrak{l}(Z_{ik})] \right) \mathtt{P}(Z_{ij} \geq \beta) + \\
\sum_{i=1}^{m+1} \sum_{j=1}^{m} \left(1 + \mathtt{E}[\mathfrak{l}(Z_{ij}) | Z_{ij} < \tfrac{1}{\beta}] + \sum_{k \neq j} \mathtt{E}[\mathfrak{l}(Z_{ik})] \right) \mathtt{P}(Z_{ij} < \tfrac{1}{\beta}).
\end{aligned} \tag{39}$$

We now estimate the terms that appear in the summation above. We have

$$\mathtt{P}(Z_{ij} \geq \beta) = \int_\beta^\infty \frac{x^{a_{ij}-1}}{\Gamma(a_{ij})} \underbrace{e^{-x}}_{\substack{= e^{-\frac{x}{2}} e^{-\frac{x}{2}} \\ \leq e^{-\frac{x}{2}} e^{-\frac{\beta}{2}}}} \mathrm{d}x \leq e^{-\frac{\beta}{2}} \underbrace{\int_\beta^\infty}_{\leq \int_0^\infty} \frac{x^{a_{ij}-1} e^{-\frac{x}{2}}}{\Gamma(a_{ij})} \mathrm{d}x \leq 2^{a_{ij}} e^{-\frac{\beta}{2}} \leq_c e^{-\frac{\beta}{2}}. \tag{40}$$

Also,

$$\mathtt{P}(Z_{ij} < \tfrac{1}{\beta}) = \int_0^{\frac{1}{\beta}} \frac{x^{a_{ij}-1}}{\Gamma(a_{ij})} \underbrace{e^{-x}}_{\leq 1} \mathrm{d}x \leq_c \frac{1}{\beta^{a_{ij}}} \leq \frac{1}{\beta}, \tag{41}$$

and the last inequality follows since $a_{ij}, \beta \geq 1$. Now,

$$\mathtt{E}[\mathfrak{l}(Z_{ij}) | Z_{ij} \geq \beta] \mathtt{P}(Z_{ij} \geq \beta) = \sum_{n=1}^{\infty} \int_{\beta^n}^{\beta^{n+1}} \mathfrak{l}(x) f_{Z_{ij}}(x) \mathrm{d}x \tag{42}$$

$$\leq \sum_{n=1}^{\infty} \int_{\beta^n}^{\beta^{n+1}} \left(\frac{x}{\log \beta} + 2 \right) f_{Z_{ij}}(x) \mathrm{d}x \tag{43}$$

$$= \int_\beta^\infty \left(\frac{x}{\log \beta} + 2 \right) f_{Z_{ij}}(x) \mathrm{d}x \leq_c e^{-\frac{\beta}{2}} \left(1 + \frac{1}{\log \beta} \right), \tag{44}$$

For the first inequality, recall that for a given $x \in [\beta^n, \beta^{n+1})$ for some $n \geq 1$, we have $\mathfrak{l}(x) = |\mathtt{a}_n|$. Since $\mathtt{a}_n = \mathtt{0}\mathtt{0}^n\mathtt{1}$, we obtain $\mathfrak{l}(x) = |\mathtt{a}_n| = n + 2 \leq \frac{\log x}{\log \beta} + 2 \leq \frac{x}{\log \beta} + 2$. The last inequality can be derived using the same arguments as in (40).

Similarly, we obtain

$$\mathtt{E}[\mathfrak{l}(Z_{ij})|Z_{ij} < \tfrac{1}{\beta}]\mathtt{P}(Z_{ij} < \tfrac{1}{\beta}) = \sum_{n=-\infty}^{-2} \int_{\beta^n}^{\beta^{n+1}} \mathfrak{l}(x) f_{Z_{ij}}(x) \mathrm{d}x \tag{45}$$

$$= \sum_{n=-\infty}^{-2} \int_{\beta^n}^{\beta^{n+1}} (|n|+1) \underbrace{f_{Z_{ij}}(x)}_{\leq_c 1} \mathrm{d}x \tag{46}$$

$$\leq_c \sum_{n=-\infty}^{-2} (|n|+1)(\beta^{n+1} - \beta^n) \tag{47}$$

$$= \frac{3\beta - 2}{\beta(\beta-1)} \leq_c \frac{1}{\beta - 1} \tag{48}$$

Also,

$$\mathtt{E}[\mathfrak{l}(Z_{ij})] = \mathtt{E}[\mathfrak{l}(Z_{ij})|Z_{ij} < \tfrac{1}{\beta}]\mathtt{P}(Z_{ij} < \tfrac{1}{\beta}) + \mathtt{E}[\mathfrak{l}(Z_{ij})|\tfrac{1}{\beta} \leq Z_{ij} < \beta]\mathtt{P}(\tfrac{1}{\beta} \leq Z_{ij} < \beta)$$
$$+ \mathtt{E}[\mathfrak{l}(Z_{ij})|Z_{ij} \geq \beta]\mathtt{P}(Z_{ij} \geq \beta) \tag{49}$$

We bound the first and the third terms using (48) and (44), respectively. Also, by the definition of $\mathfrak{l}(\cdot)$, the second conditional expectation is equal to 2. Therefore,

$$\mathtt{E}[\mathfrak{l}(Z_{ij})] \leq_c 1 + \frac{1}{\beta - 1} + e^{-\frac{\beta}{2}} \left(1 + \frac{1}{\log \beta}\right) \tag{50}$$

Applying the bounds in (40), (41), (44), (48) and (50) to (39), we obtain

$$\mathcal{R}(\mathsf{q}_\beta) - 1 \leq_c e^{-\frac{\beta}{2}} \left(1 + \frac{1}{\log \beta} + \frac{1}{\beta - 1} + e^{-\frac{\beta}{2}} \left(1 + \frac{1}{\log \beta}\right)\right) + \tag{51}$$

$$\frac{1}{\beta}\left(1 + \frac{\beta+1}{\beta-1} + e^{-\frac{\beta}{2}} \left(1 + \frac{1}{\log \beta}\right)\right). \tag{52}$$

For $\beta \geq e$, we use the bounds $e^{-\frac{\beta}{2}} \leq_c \frac{1}{\beta-1}$, $\frac{1}{\log \beta} \leq 1$, $\frac{1}{\beta} \leq \frac{1}{\beta-1}$, $\frac{\beta+1}{\beta-1} \leq_c 1$ to obtain $\mathcal{R}(\mathsf{q}_\beta) - 1 \leq_c \frac{1}{\beta-1} + \frac{1}{(\beta-1)^2} \leq_c \frac{1}{\beta-1}$. On the other hand, for $1 < \beta \leq e$, the bounds $e^{-\frac{\beta}{2}} \leq 1$, $1 \leq_c \frac{1}{\beta-1}$, $\frac{1}{\log\beta} \leq_c \frac{1}{\beta-1}$, $\frac{1}{\beta} \leq 1$, $\frac{\beta+1}{\beta-1} \leq \frac{e+1}{\beta-1}$ yield $\mathcal{R}(\mathsf{q}_\beta) - 1 \leq_c \frac{1}{\beta-1}$. This concludes the proof.

Appendix B
Proof of Proposition 1

First note that (16) is trivial if $\nu \leq \min_i \lambda_i^{-1}$. Hence, suppose that $\nu > \min_i \lambda_i^{-1}$ and let $\delta = \epsilon\sqrt{\nu}$. We have the representation $\widehat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{E}$, where each real and imaginary component of $\mathbf{E}$ has norm no greater than 1. For notational convenience, let $\Xi(\mathbf{X}, \mathbf{H}) = \det(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H})$ and note that $C(\mathbf{X}, \mathbf{H}) = \log \Xi(\mathbf{X}, \mathbf{H})$. We first rewrite $\Xi(\widehat{\mathbf{X}}, \mathbf{H})$ as

$$\Xi(\widehat{\mathbf{X}}, \mathbf{H}) = \det(\mathbf{I}_t + (\mathbf{X} + \delta\mathbf{E})(\mathbf{X}^\dagger + \delta\mathbf{E}^\dagger)\mathbf{H}^\dagger\mathbf{H}) \tag{53}$$

$$= \det(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H} + \delta \underbrace{(\mathbf{E}\mathbf{X}^\dagger + \mathbf{E}^\dagger\mathbf{X} + \delta\mathbf{E}\mathbf{E}^\dagger)}_{\triangleq \mathbf{F}} \mathbf{H}^\dagger\mathbf{H}) \quad (54)$$

$$= \Xi(\mathbf{X},\mathbf{H}) \det(\mathbf{I}_t + \delta(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{F}\mathbf{H}^\dagger\mathbf{H}) \quad (55)$$

$$= \Xi(\mathbf{X},\mathbf{H}) \det(\mathbf{I}_t + \delta\mathbf{H}^\dagger\mathbf{H}(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{F}) \quad (56)$$

Moreover, we have $(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H})^{-1} = (\mathbf{I}_t + \mathbf{U}\mathbf{N}\mathbf{U}^\dagger\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger)^{-1} = (\mathbf{I}_t + \mathbf{U}\mathbf{N}\boldsymbol{\Lambda}\mathbf{U}^\dagger)^{-1} = (\mathbf{U}\mathbf{U}^\dagger + \mathbf{U}\mathbf{N}\boldsymbol{\Lambda}\mathbf{U}^\dagger)^{-1} = \mathbf{U}(\mathbf{I}_t + \mathbf{N}\boldsymbol{\Lambda})^{-1}\mathbf{U}^\dagger$, and this gives us

$$\mathbf{H}^\dagger\mathbf{H}(\mathbf{I}_t + \mathbf{X}\mathbf{X}^\dagger\mathbf{H}^\dagger\mathbf{H})^{-1} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger\mathbf{U}(\mathbf{I}_t + \mathbf{N}\boldsymbol{\Lambda})^{-1}\mathbf{U}^\dagger = \mathbf{U}\underbrace{\boldsymbol{\Lambda}(\mathbf{I}_t + \mathbf{N}\boldsymbol{\Lambda})^{-1}}_{\triangleq \mathbf{Z}}\mathbf{U}^\dagger, \quad (57)$$

and $\Xi(\widehat{\mathbf{X}},\mathbf{H}) = \Xi(\mathbf{X},\mathbf{H}) \det(\mathbf{I}_t + \delta\mathbf{U}\mathbf{Z}\mathbf{U}^\dagger\mathbf{F})$. Now, let $\mathbf{W} = [w_{ij}]_{t\times t} = \delta\mathbf{U}\mathbf{Z}\mathbf{U}^\dagger\mathbf{F}$. Let $\omega_1, \ldots, \omega_t$ denote the eigenvalues of $\mathbf{W}$, and $D(a,r) = \{z \in \mathbb{Z} : |z-a| \le r\}$ denote the closed disk centered at a with radius r. According to Gershgorin circle theorem, we have

$$\{\omega_1,\ldots,\omega_t\} \subset \bigcup_{i=1}^{t} D\left(w_{ii}, \sum_{j\neq i} |w_{ij}|\right) \subset \bigcup_{i=1}^{t} D\left(0, |w_{ii}| + \sum_{j\neq i} |w_{ij}|\right) \subset D(0, \|\mathbf{W}\|_1), \quad (58)$$

where for any complex matrix $\mathbf{A} = [a_{ij}]_{M\times N}$, we let $\|\mathbf{A}\|_1 \triangleq \sum_{i=1}^{M}\sum_{j=1}^{N} |a_{ij}|$. Moreover, since $\mathbf{W}$ is a Hermitian matrix, its eigenvalues are real. Therefore, $-\|\mathbf{W}\|_1 \le \omega_1, \ldots, \omega_t \le \|\mathbf{W}\|_1$, and

$$\Xi(\widehat{\mathbf{X}},\mathbf{H}) = \Xi(\mathbf{X},\mathbf{H}) \prod_{i=1}^{t} (1+\omega_i) \ge \Xi(\mathbf{X},\mathbf{H})[(1 - \|\mathbf{W}\|_1)^+]^t. \quad (59)$$

We now estimate $\|\mathbf{W}\|_1$. First, since $\|\cdot\|_1$ is a sub-multiplicative norm, we have

$$\|\mathbf{W}\|_1 \le \delta\|\mathbf{U}\|_1\|\mathbf{Z}\|_1\|\mathbf{U}^\dagger\|_1\|\mathbf{F}\|_1 = \delta\|\mathbf{U}\|_1^2\|\mathbf{Z}\|_1\|\mathbf{F}\|_1. \quad (60)$$

We now find upper bounds on each of the $\|\cdot\|_1$ terms. For any complex matrix $\mathbf{A} = [a_{ij}]_{M\times N}$, it can be shown using Hölder's inequality that $\|\mathbf{A}\|_1 \le \sqrt{MN}\|\mathbf{A}\|$. In particular, $\|\mathbf{U}\|_1^2 \le t^2\|\mathbf{U}\| = t^2$. Regarding $\|\mathbf{Z}\|_1$, note that

$$\mathbf{Z} = \operatorname{diag}\left(\frac{\lambda_1}{1 + (\nu\lambda_1 - 1)^+}, \ldots, \frac{\lambda_m}{1 + (\nu\lambda_m - 1)^+}, 0, \ldots, 0\right). \quad (61)$$

For any $i \in \{1,\ldots,m\}$, if $\nu \ge \frac{1}{\lambda_i}$, then $\frac{\lambda_i}{1+(\nu\lambda_i-1)^+} = \frac{1}{\nu}$, and otherwise if $\nu < \frac{1}{\lambda_i}$, then $\frac{\lambda_i}{1+(\nu\lambda_i-1)^+} = \lambda_i < \frac{1}{\nu}$. In either case, $\frac{\lambda_i}{1+(\nu\lambda_i-1)^+} \le \frac{1}{\nu}$, and therefore we have the estimate $\|\mathbf{Z}\|_1 \le \frac{t}{\nu}$. For $\|\mathbf{F}\|_1$, we obtain

$$\|\mathbf{F}\|_1 = \|\mathbf{E}\mathbf{X}^\dagger + \mathbf{E}^\dagger\mathbf{X} + \delta\mathbf{E}\mathbf{E}^\dagger\|_1 \le \|\mathbf{E}\|_1\|\mathbf{X}^\dagger\|_1 + \|\mathbf{E}^\dagger\|_1\|\mathbf{X}\|_1 + \delta\|\mathbf{E}\|_1\|\mathbf{E}^\dagger\|_1$$

$$= 2\|\mathbf{E}\|_1\|\mathbf{X}\|_1 + \delta\|\mathbf{E}\|_1^2 \leq 2t^2\|\mathbf{X}\|_1 + \delta t^4 \leq 2t^3\|\mathbf{X}\| + \delta t^4 = 2t^3\sqrt{\sum_{i=1}^{m}\left(\nu - \frac{1}{\lambda_m}\right)^+} + \delta t^4$$

$$\leq 2t^3\sqrt{\nu m} + \delta t^4 = \sqrt{\nu}(2t^3\sqrt{m} + \epsilon t^4) \leq 3t^4\sqrt{\nu} \quad (62)$$

Combining the estimates on $\|\mathbf{U}\|_1$, $\|\mathbf{X}\|_1$, and $\|\mathbf{F}\|_1$, we obtain $\|\mathbf{W}\|_1 \leq 3t^7\epsilon$. For $\epsilon \leq \frac{1}{6t^8}$, we have $\|\mathbf{W}\|_1 \leq 1$, and therefore, according to (59),

$$\Xi(\widehat{\mathbf{X}}, \mathbf{H}) \geq \Xi(\mathbf{X}, \mathbf{H})(1 - \|\mathbf{W}\|_1)^t \geq \Xi(\mathbf{X}, \mathbf{H})(1 - 3t^7\epsilon)^t \geq \Xi(\mathbf{X}, \mathbf{H})(1 - 3t^8\epsilon). \quad (63)$$

Taking logarithms, and applying the bound $\log(1 - x) \geq -2x$, $0 \leq x \leq \frac{1}{2}$, we obtain (16).

Appendix C
Proof of Proposition 3

Let $\lambda'_1, \ldots, \lambda'_m \in \mathbb{R}_{>0}$ and $\omega \geq 1$. Let μ' be defined as in (14) with respect to the eigenvalues $\lambda'_1, \ldots, \lambda'_m$, i.e. $\mu' = (\frac{2^R}{\prod_{i\in\mathcal{I}'}\lambda'_i})^{1/|\mathcal{I}'|}$ with $i \in \mathcal{I}' \iff \mu' \geq \frac{1}{\lambda'_i}$ for every $i \in \{1, \ldots, m\}$. It is sufficient to prove that for any $\mathbf{H}$ with

$$\lambda'_i \overset{(64a)}{\leq} \lambda_i \overset{(64b)}{\leq} \omega\lambda'_i, \forall i \in \{1, \ldots, m\}, \quad (64)$$

we have

$$\frac{\mu'}{\omega} \overset{(65a)}{\leq} \mu_\star \overset{(65b)}{\leq} \mu'. \quad (65)$$

To prove (65a) and (65b), we first note the inequalities

$$\frac{1}{\lambda_j} \overset{(66a)}{\geq} \mu_\star = \left(\frac{e^d}{\prod_{i\in\mathcal{I}_\star}\lambda_i}\right)^{\frac{1}{|\mathcal{I}_\star|}} \overset{(66b)}{\geq} \frac{1}{\lambda_k}, \forall k \in \mathcal{I}_\star, \forall j \in \mathcal{I}' - \mathcal{I}_\star, \quad (66)$$

and

$$\frac{1}{\lambda'_j} \overset{(67a)}{\geq} \mu' = \left(\frac{e^d}{\prod_{i\in\mathcal{I}'}\lambda'_i}\right)^{\frac{1}{|\mathcal{I}'|}} \overset{(67b)}{\geq} \frac{1}{\lambda'_k}, \forall k \in \mathcal{I}', \forall j \in \mathcal{I}_\star - \mathcal{I}'. \quad (67)$$

Suppose $\mathcal{I}' - \mathcal{I}_\star \neq \emptyset$. Then, there is an index $\ell \in \mathcal{I}' - \mathcal{I}_\star$ and obviously also $\ell \in \mathcal{I}'$. These give us $\mu' \overset{(67b)}{\geq} \frac{1}{\lambda'_\ell} \overset{(64a)}{\geq} \frac{1}{\lambda_\ell} \overset{(66a)}{\geq} \mu_\star$, and proves (65b) when $\mathcal{I}' - \mathcal{I}_\star \neq \emptyset$. If further $\mathcal{I}_\star - \mathcal{I}' \neq \emptyset$, then, there exists $\ell \in \mathcal{I}_\star - \mathcal{I}'$ such that $\mu_\star \overset{(66b)}{\geq} \frac{1}{\lambda_j} \overset{(64b)}{\geq} \frac{1}{\omega\lambda'_j} \overset{(67a)}{\geq} \frac{\mu'}{\omega}$. This proves (65a) when $\mathcal{I}' - \mathcal{I}_\star \neq \emptyset$ and $\mathcal{I}_\star - \mathcal{I}' \neq \emptyset$. Now suppose that $\mathcal{I}' - \mathcal{I}_\star \neq \emptyset$ and $\mathcal{I}_\star - \mathcal{I}' = \emptyset$. We have $\mathcal{I}' = (\mathcal{I}' - \mathcal{I}_\star) \cup \mathcal{I}_\star$, and therefore,

$$\frac{\mu_\star}{\mu'} = e^{d(\frac{1}{|\mathcal{I}_\star|} - \frac{1}{|\mathcal{I}'|})}\frac{(\prod_{i\in\mathcal{I}'}\lambda'_i)^{\frac{1}{|\mathcal{I}'|}}}{(\prod_{i\in\mathcal{I}_\star}\lambda_i)^{\frac{1}{|\mathcal{I}_\star|}}} = e^{d(\frac{1}{|\mathcal{I}_\star|} - \frac{1}{|\mathcal{I}'|})}\frac{(\prod_{i\in\mathcal{I}'-\mathcal{I}_\star}\lambda'_i)^{\frac{1}{|\mathcal{I}'|}}(\prod_{i\in\mathcal{I}_\star}\lambda'_i)^{\frac{1}{|\mathcal{I}'|}}}{(\prod_{i\in\mathcal{I}_\star}\lambda_i)^{\frac{1}{|\mathcal{I}_\star|}}}. \quad (68)$$

On the other hand, according to (67b), we have

$$\left(\frac{e^d}{\prod_{i\in\mathcal{I}'}\lambda'_i}\right)^{\frac{1}{|\mathcal{I}'|}} \geq \frac{1}{\lambda'_k},\ \forall k\in\mathcal{I}'-\mathcal{I}_\star. \tag{69}$$

Multiplying all these $|\mathcal{I}'-\mathcal{I}_\star| = |\mathcal{I}'|-|\mathcal{I}_\star|$ inequalities, we obtain

$$\left(\frac{e^d}{\prod_{i\in\mathcal{I}'}\lambda'_i}\right)^{\frac{|\mathcal{I}'|-|\mathcal{I}_\star|}{|\mathcal{I}'|}} \geq \frac{1}{\prod_{k\in\mathcal{I}'-\mathcal{I}_\star}\lambda'_k}, \tag{70}$$

and, equivalently,

$$\left(\prod_{i\in\mathcal{I}'-\mathcal{I}_\star}\lambda'_i\right)^{\frac{1}{|\mathcal{I}'|}} \geq \left(\frac{\prod_{i\in\mathcal{I}_\star}\lambda'_i}{e^d}\right)^{\frac{1}{|\mathcal{I}_\star|}-\frac{1}{|\mathcal{I}'|}} \tag{71}$$

Substituting (71) to (68), and applying (64b) to the denominator of (68), we have $\mu_\star \geq \frac{\mu'}{\omega}$. This concludes the proof of the lemma for the case $\mathcal{I}'-\mathcal{I}_\star \neq \emptyset$.

Now, suppose $\mathcal{I}'-\mathcal{I}_\star = \emptyset$. If $\mathcal{I}'=\mathcal{I}_\star$, we have

$$\frac{\mu'}{\omega} = \left(\frac{e^d}{\prod_{i\in\mathcal{I}_\star}(\omega\lambda'_i)}\right)^{\frac{1}{|\mathcal{I}_\star|}} \overset{(64b)}{\leq} \left(\frac{e^d}{\prod_{i\in\mathcal{I}_\star}\lambda_i}\right)^{\frac{1}{|\mathcal{I}_\star|}} = \mu_\star \overset{(64a)}{\leq} \left(\frac{e^d}{\prod_{i\in\mathcal{I}_\star}\lambda'_i}\right)^{\frac{1}{|\mathcal{I}_\star|}} = \mu', \tag{72}$$

and thus (65a) and (65b) are easily proved. What is left is thus the case $\mathcal{I}'-\mathcal{I}_\star = \emptyset$ with $\mathcal{I}_\star-\mathcal{I}' \neq \emptyset$. First, note that there is an index $\ell\in\mathcal{I}_\star-\mathcal{I}'$ such that $\mu_\star \overset{(66b)}{\geq} \frac{1}{\lambda_\ell} \overset{(64b)}{\geq} \frac{1}{\omega\lambda'_\ell} \overset{(67a)}{\geq} \frac{\mu'}{\omega}$, and this proves (65a). For (65b), starting from the fraction $\frac{\mu'}{\mu_\star}$, we can go through the same arguments as in (68) to find out that $\frac{\mu'}{\mu_\star} \geq 1$. This concludes the proof.

References

[1] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block fading channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1273–1289, May 2001.

[2] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[3] C. W. Sung and W. S. Wong, "A distributed fixed-step power control algorithm with quantization and active link quality protection," *IEEE Trans. Veh. Tech.*, vol. 48, no. 2, Mar. 1999.

[4] K. K. Leung, C. W. Sung, W. S. Wong, and T. M. Lok, "Convergence theorem for a general class of power-control algorithms," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1566–1574, Sept. 2004.

[5] H. Su, E. Geraniotis, "Adaptive closed-loop power control with quantized feedback and loop filtering," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 76–86, Jan. 2002.

[6] B. Khoshnevis and W. Yu, "Joint power control and beamforming codebook design for MISO channels under the outage criterion," *J. of Franklin Institute Special Issue on Modelling and Simulation in Advanced Commun.*, vol. 349, no. 1, pp. 140–161, Feb. 2012. (invited).

[7] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 785–798, May. 2002.

[8] T. T. Kim and M. Skoglund, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2743–2759, Aug. 2007.

[9] A. Khoshnevis and A. Sabharwal, "On the asymptotic performance of multiple antenna channels with quantized feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3869–3877, Oct. 2008.

[10] N. Ahmed, M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Outage minimization with limited feedback for the fading relay channel," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 659–666, Apr. 2006.

[11] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[12] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol 10, no. 6, pp. 585–595, Dec. 1999.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] B. Mondal and R. W. Heath Jr., "Performance analysis of quantized beamforming MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4753–4766, Dec. 2006.

[21] 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.

[22] J. Zheng and B. D. Rao, "Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4612–4626, Sept. 2007.

[23] 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.

[24] 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.

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

[26] ——, "Capacity of a multi-antenna fading channel with a quantized precoding matrix," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1218–1234, Mar. 2009.

[27] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.

[28] R. T. Krishnamachari, M. K. Varanasi, and K. Mohanty, "MIMO systems with quantized covariance feedback," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 485–495, Jan. 2014.

[29] 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.

[30] 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.

[31] 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.

[32] V. Raghavan, V. Veeravalli, "Ensemble properties of RVQ-based limited-feedback beamforming codebooks", *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8224–8249, Dec. 2013.

[33] R. T. Krishnamachari and M. K. Varanasi, "Finite-rate feedback of input covariance matrices in MIMO systems," *IEEE Intl. Symp. on Inf. Theory*, June 2010.

[34] ——, "MIMO performance under covariance matrix feedback," *IEEE Intl. Symp. on Inf. Theory*, Aug. 2011.

[35] Y. Isukapalli, R. Annavajjala, and B. D. Rao, "Average SEP and BEP analysis of transmit beamforming for MISO systems with imperfect feedback and M-PSK constellation," *IEEE Int. Symp. on Personnel, Indoor, Mobile Radio Commun.*, Sept. 2007.

[36] 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.

[37] 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.

[38] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[39] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, July 2007.

[40] 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.

[41] 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.

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

[43] E. Koyuncu and H. Jafarkhani, "Variable-length limited feedback beamforming in multiple-antenna fading channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7140–7165, Nov. 2014.

[44] X. Liu, E. Koyuncu, and H. Jafarkhani, "Multicast networks with variable-length limited feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 252–264, Jan. 2015.

[45] ——, "Cooperative quantization for two-user interference channels," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2698–2712, June 2015.

[46] N. Prasad and M. K. Varanasi, "MIMO outage capacity in the high SNR regime," *IEEE Intl. Symp. Inf. Theory*, Sept. 2005.