On the Information Rate of MIMO Systems With Finite Rate Channel State Feedback Using Beamforming and Power On/Off Strategy
Wei Dai, Member, IEEE, Youjian (Eugene) Liu, Member, IEEE, Brian Rider, and Vincent K. N. Lau, Senior Member, IEEE

Abstract—It is well known that multiple-input multiple-output (MIMO) systems have high spectral efficiency, especially when channel state information at the transmitter (CSIT) is available. In many practical systems, it is reasonable to assume that the CSIT is obtained by a limited (i.e., finite rate) feedback and is therefore imperfect. We consider the design problem of how to use the limited feedback resource to maximize the achievable information rate. In particular, we develop a low complexity power on/off strategy with beamforming (or Grassmann precoding), and analytically characterize its performance. Given the eigenvalue decomposition of the covariance matrix of the transmitted signal, refer to the eigenvectors as beams, and to the corresponding eigenvalues as the beam’s power. A power on/off strategy means that a beam is either turned on with a constant power or turned off. We will first assume that the beams match the channel perfectly and show that the ratio between the optimal number of beams turned on and the number of antennas converges to a constant when the numbers of transmit and receive antennas approach infinity proportionally. This motivates our power on/off strategy where the number of beams turned on is independent of channel realizations but is a function of the signal-to-noise ratio (SNR). When the feedback rate is finite, beamforming cannot be perfect, and we characterize the effect of imperfect beamforming by quantization bounds on the Grassmann manifold. By combining the results for power on/off and beamforming, a good approximation to the achievable information rate is derived. Simulations show that the proposed strategy is near optimal and the performance approximation is accurate for all experimented SNRs.

Index Terms—Beamforming, Grassmann manifold, limited feedback, multiple-input multiple-output (MIMO), power on/off, precoding.

Manuscript received September 01, 2005; revised April 01, 2009. Current version published October 21, 2009. This work was supported in part by NSF under Grants CCF-0728955, ECCS-0725915, DMS-0505680, and DMS-0645756. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia, September 2005.

W. Dai is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: weidai07@uiuc.edu).

Y. (E.) Liu is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309 USA (e-mail: eugeneliu@ieee.org).

B. Rider is with the Mathematics Department, University of Colorado, Boulder, CO 80309 USA (e-mail: brian.rider@colorado.edu).

V. K. N. Lau is with the Department of Electrical and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: eeknlau@ee.ust.hk).

Communicated by G. Taricco, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2009.2030470

I. INTRODUCTION

This paper considers multiple-input multiple-output (MIMO) systems with finite rate channel state feedback. Multiple-antenna wireless communication systems, also known as MIMO systems, have high-spectral efficiency. The full potential of MIMO systems relies on channel state information (CSI); when perfect CSI is available at both the transmitter and receiver (CSITR), the MIMO channel can be transformed into a set of parallel subchannels; the optimal transmission power allocation on each subchannel follows the water filling principle [1]. While CSI at the receiver side (CSIR) can be obtained via channel training and estimation processes, CSI at the transmitter side (CSIT) is usually obtained from feedback from the receiver. Clearly, perfect CSIT requires infinite feedback rates, which is not practical. On the other hand, in practical systems, such as Universal Mobile Telecommunications System—High Speed Downlink Packet Access (UMTS-HSDPA) [2], there is a control field which can be used to carry a certain number of channel state feedback bits on a per-fading block basis. It is therefore important to consider MIMO systems with finite rate channel state feedback.

For a given feedback rate, this paper addresses two basic questions: how much benefit can feedback provide and how to achieve that benefit. In order to achieve and compute the information rate for a given feedback rate, a joint design of transmission and feedback strategies is required. This results in the joint optimization problem stated in [3] and [4] (see Section II-A as well), which is extremely difficult to solve. For memoryless channels, it has been shown that the information theoretic limit can be achieved by memoryless transmission and feedback strategies [3], [4]. However, the explicit forms of the optimal strategies are still unknown; in [3] and [4], a Lloyd algorithm is employed to obtain suboptimal numerical solution.

The joint optimization problem may be simplified if the transmission is restricted to follow a power on/off strategy as described below. In a general setting, the optimal transmission strategy is to choose the covariance matrix of the transmitted Gaussian coded symbols according to the current feedback [3], [4]. By the eigenvalue decomposition, the covariance matrix can be decomposed to a unitary matrix and a nonnegative diagonal matrix, which are called beamforming matrix (also known as precoding matrix), and power control matrix, respectively. We refer to each column vector of the beamforming matrix as a beam and the diagonal element corresponding to a beam as the power on that beam. A power on/off strategy means that a beam is either turned on with a constant power or turned off. As we
will show later, this assumption significantly simplifies the design and analysis without greatly sacrificing performance. It has been already shown in [5] and [6] that power on/off is near optimal for single antenna fading channels and parallel Gaussian channels respectively. We shall empirically demonstrate that the same conclusion holds for MIMO channels as well.

The main contribution of this paper is to develop a low complexity power on/off strategy and analytically characterize its performance. The original joint optimization problem is decoupled into two individual optimization problems: the first one is to find the optimal number of beams to turn on (henceforth, on-beams), which is related to power control; the second one is to choose the beamforming matrices according to the channel realizations (referred to as beamforming for short). By analyzing the effects of power control and beamforming, we are able to approximately characterize the achievable information rate.

To isolate the effect of imperfect beamforming, we first study a power on/off strategy with perfect beamforming. We assume an artificial scenario where the feedback rate is infinite so that beamforming at transmitter perfectly decomposes the MIMO channel into independent subchannels. An asymptotic analysis, in which the numbers of the transmit and receive antennas approach infinity simultaneously, leads to the following results.

- Define the minimum number of transmit and receive antennas as the dimension of a MIMO system. We prove that the ratio between the optimal number of on-beams and the system dimension converges to a constant almost surely. This suggests a power on/off strategy with a constant number of on-beams, where the number of on-beams is independent of channel realizations but is a function of the operating signal-to-noise-ratio (SNR). (The assumption of a constant number of on-beams is crucial to analyze the effect of imperfect beamforming later.)
- We show that the optimal ratio between the number of on-beams and the system dimension is an increasing function of SNR.
- We derive a criterion to find this optimal ratio and asymptotic formulas to calculate its value and the corresponding information rate.
- At an empirical level we demonstrate that power on/off strategy is near optimal for MIMO systems by comparing it with power water filling.

Assuming the number of on-beams is constant, the effect of imperfect beamforming has been considered previously. As a special case of general beamforming, analysis can be significantly simplified by assuming either single receive antenna [7]–[11] or single on-beam [11], [12]. Multiple transmit antenna selection was studied in [13], [14]. Note that antenna selection is equivalent to choosing the beamforming matrices to be truncated identity matrices. Restrictions on the number of on-beams or structures of beamforming matrices significantly sacrifice the full potential of MIMO systems. For general beamforming, criteria and algorithms for codebook design were studied in [15]–[17]. The effect of channel quantization was analyzed in [18], [19] and then refined in [20]. However, the analysis was based on Barg-Nogin’s formula [21], which is valid only when the number of receive antennas is fixed and the number of transmit antennas approaches infinity. When the numbers of receive and transmit antennas are in the same order, performance at high SNRs was analyzed in [22] while the general SNR regime was treated in [23], [24].

We introduce a single parameter, termed the power efficiency factor, to quantify the effect of imperfect beamforming. By resorting to our closed-form formulas for quantization bounds on the Grassmann manifold [25], we tightly bound the power efficiency factor as a function of feedback rate. As a result, the information rate of the proposed power on/off strategy can be well characterized.

Our approach exhibits the following advantages.

- The performance analysis is valid for all SNR regimes, showing in particular that the feedback gain is more significant for low and median SNRs. In general the results provide a guide for measuring the effect of feedback.
- While our theoretical approximation is based on asymptotics, simulations show it remains accurate for systems with small numbers of receive and transmit antennas. This may be understood as a familiar concentration of measure phenomenon connected with random matrix theory.
- The proposed suboptimal strategy is empirically near-optimal. To show that, we design a general power on/off strategy in Section IV, where the number of on-beams is not fixed. Numerical comparison between the proposed and the general power on/off strategies suggests the near-optimality of the proposed strategy.

This paper is organized as follows. The system model and the related design problem are outlined in Section II, where preliminary knowledge about random matrices and Stiefel and Grassmann manifolds are also presented. In Section III, the power on/off strategy with a constant number of on-beams is derived as the asymptotically optimal solution for perfect beamforming. Section IV considers the effect of imperfect beamforming due to finite rate channel state feedback. Section V demonstrates the near optimality of fixing number of on-beams. Conclusions are given in Section VI.

II. PRELIMINARIES

In this section, we describe the system model and present some needed facts about random matrices and Stiefel and Grassmann manifolds.

Throughout, we use $\mathbb{Z}^+$ to denote the set of positive integers, $\mathbb{R}^k$ and $\mathbb{C}^k$ for the $k$-dimensional real and complex vector spaces respectively, and $\mathbb{C}^{k\times l}$ for the space of $k \times l$ complex matrices. Also, $\mathbf{I}_k$ denotes the $k \times k$ identity matrix, $\mathbf{A}^\dagger$ the conjugate transpose of a matrix $\mathbf{A}$, $\operatorname{tr} (\cdot)$ the usual matrix trace, and $\operatorname{rank} (\cdot)$ the rank of a matrix. We use $\| \cdot \|_F$ for the (matrix) Frobenius norm, and $\| \cdot \|$ to denote the determinant of a matrix or the cardinality of a set according to its context. $E_X [\cdot]$ stands for expectation with respect to the random variable $X$, and $\arg \max$ or $\arg \min$ denote the global maximizer or minimizer of a given quantity.

A. System Model and the Corresponding Design Problem

A communication system with $L_T$-transmit antennas and $L_R$-receive antennas is depicted in Fig. 1. Let $\mathbf{t} \in \mathbb{C}^{L_T}$ be the transmitted signal, $\mathbf{y} \in \mathbb{C}^{L_R}$ be the received signal, $\mathbf{H} \in \mathbb{C}^{L_R \times L_T}$ be the channel state matrix and $\mathbf{z} \in \mathbb{C}^{L_R}$ be the
additive Gaussian noise with zero mean and unit covariance matrix $E[zz^\dagger] = I_{LR}$. Then the received signal is modeled as

$$y = Ht + z. \quad (1)$$

In this paper, the following Rayleigh block fading channel is considered. The entries of $H$ are independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian variables with zero mean and unit variance ($CN(0, 1)$) and $H$ is i.i.d. from one block to another. During each fading block, the channel remains unchanged. At the beginning of each channel block, the channel state $H$ is assumed to be perfectly estimated at the receiver, then quantized to finite bits and fed back to the transmitter through a feedback channel. The time required for channel estimation and feedback can be ignored when the length of a fading block is long. The feedback channel is assumed to be error-free and the introduced delay can also be ignored. The rate of the feedback is limited, up to $R_b$ bits/channel realization for some small $R_b$, $\in \mathbb{Z}^+$. After receiving the channel state feedback, the transmitter transmits the encoded signal $t$ according to the current feedback. A finite alphabet model has been studied in [3]. It has been shown that memoryless transmission and feedback strategies can achieve the information theoretic limit.

Ideally, one would like to jointly design transmission and feedback strategies to maximize the achievable information rate. Given a feedback strategy, the optimal transmitted signal is circular symmetric Gaussian distributed with zero mean and covariance matrix adapted to the current channel feedback [3]. Denote the covariance matrix of the transmitted signal by $\Sigma = E[tt^\dagger]$. Denote the codebook of the covariance matrices by

$$B_\Sigma = \{ \Sigma_i \in \mathbb{C}^{LT \times LR} : \ 1 \leq i \leq 2^{R_B} \}. \quad (2)$$

The feedback function $\varphi(\cdot)$ is a mapping from the space of $H$ to the index set $\{ i : 1 \leq i \leq 2^{R_B}\}$. The corresponding optimization problem is to find the optimal codebook $B_\Sigma$ and the optimal feedback function $\varphi(\cdot)$ to maximize the information rate

$$\max_{B_\Sigma, \varphi(\cdot)} E[H][\log |I + H\Sigma_{\varphi(H)}H^\dagger|] \quad (3)$$

with the average power constraint

$$E[H][\text{tr}(\Sigma_{\varphi(H)})] \leq \rho. \quad (4)$$

As demonstrated in [3], this joint optimization problem is difficult to solve.

To make the analysis tractable, we resort to the following suboptimal power on/off strategy. Denote the eigenvalue decomposition of the covariance matrix by $\Sigma = QQ^\dagger$ where the matrices $Q$ and $P$ are called beamforming matrix and power control matrix, respectively. Refer to the column vectors of $Q$ as beams. Term the beam corresponding to a positive transmission power (positive singular value of $\Sigma$) an on-beam. The statistics of the transmitted signal is uniquely determined by the on-beams and the corresponding powers. In our power on/off model, every on-beam corresponds to a constant power $P_{on}$. With a slight abuse of notation, let $Q$ be the beamforming matrix only containing the on-beams. The transmitted Gaussian signal $t$ is then

$$t = Qx$$

where $x \in \mathbb{C}^{s \times 1}$ is a random Gaussian vector with zero mean and covariance matrix $P_{on}I_s$, $s$ is the number of on-beams, and $Q \in \mathbb{C}^{LT \times s}$ satisfies $Q^\dagger Q = I_s$. Hence, the signal model for power on/off strategy is given by

$$y = HQx + z. \quad (5)$$

Note that the number of on-beams $s$ is the rank of the beamforming matrix $Q$. The feedback only needs to specify $Q$. Denote the codebook of beamforming matrices by

$$B = \{ Q_i \in \mathbb{C}^{LT \times s} : Q_i^\dagger Q_i = I_s, \ 0 \leq s \leq L_T, \ 1 \leq i \leq 2^{R_B} \}. \quad (6)$$

The original optimization problem (3) is then reduced to one of the following form.

**Problem 1 (Power On/Off Strategy Design Problem):** Find the beamforming codebook $B$, feedback function $\varphi(\cdot)$ and $P_{on}$ to maximize the information rate

$$\max_{P_{on}, B_\varphi(t)} E[H][\log |I + P_{on}HQ_{\varphi(H)}Q_{\varphi(H)}^\dagger|H^\dagger] \quad (7)$$

with the average power constraint

$$E[H][\text{tr}(Q_{\varphi(H)}Q_{\varphi(H)}^\dagger)] = P_{on}E[H][s] \leq \rho \quad (8)$$

where the number of on-beams $s = s(H) = \text{rank}(Q_{\varphi(H)})$ is a function of the current channel realization $H$.

As we will show later, the power on/off assumption is the key to decouple the beamforming codebook design and feedback function design.

### B. Random Matrix Theory

We review relevant results on the spectra of large random matrices. Recall that $H$ is an $L_R \times L_T$ random matrix with i.i.d. complex Gaussian entries of zero mean and unit variance. Define $m \triangleq \min\{L_T, L_R\}$, $n \triangleq \max\{L_T, L_R\}$, and

$$W \triangleq \begin{cases} \frac{1}{m} HH^\dagger, & \text{if } L_R < L_T \\ \frac{1}{m} H^\dagger H, & \text{if } L_R \geq L_T. \end{cases}$$

Let $\{\lambda_j\}$ be the set of the eigenvalues of $W$. Introduce the empirical eigenvalue distribution of $W$

$$F_m(\lambda) \triangleq \frac{1}{m} \left| \{ j : \lambda_j < \lambda \} \right|. \quad (9)$$
Then, as $m,n \to \infty$ with $\frac{n}{m} \to \tau$ (or $L_T/L_R \to 1/\tau$),
\[
\lim_{(n,m) \to \infty} \frac{1}{m} \sum_{i=1}^m g(\lambda_i) = \frac{1}{\pi \tau} \int g(\lambda) \, d\mu_{\lambda} \approx \frac{1}{2\pi \tau} \sqrt{\frac{\lambda^+ - \lambda^-}{\lambda - \lambda^{-}}} \, : \, d\lambda
\] 
holds for all $A \in S_{LT,LT}(\mathbb{C})$, $B \in S_{s,s}(\mathbb{C})$ and measurable set $M$ [31]. Here, $A \mathcal{M} = \{AC : C \in M\}$ if $M \subset S_{LT,s}(\mathbb{C})$, $A \mathcal{M} = \{P(AC) : P(C) \in \mathcal{M}\}$ if $M \subset G_{LT,s}(\mathbb{C})$ where $P(C)$ denotes the plane generated by $C$, and $\mathcal{M} \mathcal{B}$ is similarly defined. The invariant probability measure defines the isotropic distribution on $S_{LT,s}(\mathbb{C})$ or $G_{LT,s}(\mathbb{C})$ [31]. It is also true that if $A \in S_{LT,s}(\mathbb{C})$ is isotropically distributed, so is the generated plane $P(A) \in G_{LT,s}(\mathbb{C})$ [32].

### III. Power On/Off Strategies With Perfect CSIT

To isolate the effect of power on/off and imperfect beamforming, this section studies power on/off with perfect beamforming. Note that perfect beamforming requires infinite feedback rate. The effect of imperfect beamforming introduced by finite rate feedback will be treated in Section IV.

Henceforth, we let $m = \min(L_T,L_R)$, $n = \max(L_T,L_R)$, and $\mathbf{W} = \frac{1}{m} \mathbf{HH}^\dagger$ if $L_R < L_T$ or $\mathbf{W} = \frac{1}{m} \mathbf{H}^\dagger \mathbf{H}$ if $L_R \geq L_T$. Denote the $i$th largest eigenvalue of $\mathbf{W}$ by $\lambda_i$.

Section III-A describes the corresponding optimization problem, Section III-B solves the optimization problem in the limit as $L_T$ and $L_R$ approach infinity simultaneously, and Section III-C shows that this asymptotic solution is near optimal for MIMO systems with finite many antennas.

#### A. The Design Problem With Perfect Beamforming

When perfect CSIT is available, the optimal beamforming matrix is obtained as follows.

**Lemma 2:** Given $P_{an} \in \mathbb{R}^+$, $\mathbf{H} \in \mathbb{C}^{LT \times LR}$ and $1 \leq s \leq LT$,

\[
\max_{Q \in S_{LT,s}(\mathbb{C})} \log |\mathbf{I}_{LR} + P_{an} \mathbf{HQ}^\dagger \mathbf{Q}^\dagger| = \sum_{j=1}^s \log(1 + mP_{an} \lambda_j)
\]

and it is achieved by $Q = V_s$ where $V_s$ contains the $s$ right singular vectors of $\mathbf{H}$ corresponding to the $s$ largest singular values of $\mathbf{H}$.

This lemma is proved in Appendix A and points to a simplification of the corresponding design problem.

**Problem 2 (Power On/Off Design With Perfect Beamforming):** Find the optimal on-beam counting function $s : \mathbb{H} \mapsto \mathbb{Z}^+$ function and $P_{an}$ to maximize the information rate

\[
\max_{P_{an},s(\cdot)} E_{\mathbf{H}} \left[ \sum_{i=1}^s \log(1 + mP_{an} \lambda_i) \right]
\]

subject to $E_{\mathbf{H}}[sP_{an}] \leq \rho$.

**Theorem 1:** The optimal $s(\mathbf{H})$ function defined in Problem 2 is given by

\[
s(\mathbf{H}) = \left\{ k : \lambda_k \geq \kappa \right\}
\]

where $\kappa \in \mathbb{R}^+$ is chosen to satisfy the average power constraint

\[
P_{an} E_{\mathbf{H}}[s(\mathbf{H})] = \rho.
\]
Proof: See Appendix B.

The intuition behind the proof is that all the “good” beams (corresponding to $\lambda \geq \kappa$) and only the “good” beams should be turned on. This same intuition will guide the proof of Theorem 6 later on.

Although the form of the optimal $s(H)$ function has been found, it remains difficult to determine the key parameters (the optimal $P_{\text{on}}$ and $\kappa$) and thus the corresponding information rate $T$. In contrast to the water filling solution in which the Lagrange multiplier is uniquely determined by $p[1]$, power on/off strategy has uncountably many pairs of $P_{\text{on}}$ and $\kappa$ for a given $\rho$. Monte Carlo simulations might be used to find the optimal $P_{\text{on}}$ and $\kappa$.

On the other hand, as we will show next, if one lets the number of transmit and receive antennas approach infinity simultaneously, an asymptotic analysis produces the needed quantities.

B. MIMO Systems With Infinitely Many Antennas

Now consider the situation where the numbers of transmit and receive antennas approach infinity simultaneously. In this asymptotic regime the power on/off strategy simplifies to the point that we may efficiently compute the various key parameters introduced just above. In addition, we also derive asymptotic formulas for the performance in the CSITR case. To the best of the authors’ knowledge, these asymptotic formulas have not been presented previously.

Define the normalized number of on-beams by $\bar{s}(H) \triangleq \frac{1}{m} s(H)$, the normalized on-power by $\bar{P}_{\text{on}}(H) = m P_{\text{on}}(H)$, and the normalized information rate by $\bar{T}(H) = \frac{1}{m} T(H)$. The chief result in this direction is the almost sure convergence of $\bar{s}(H)$ and $\bar{T}(H)$.

Theorem 2: For a given SNR $\rho$, let $L_T, L_R \rightarrow \infty$ with $\frac{n}{m} \rightarrow \tau (L_T / L_R \rightarrow \tau$ or $1 / \tau)$. Then

$$\lim_{(n,m) \rightarrow \infty} \bar{s}(H) = \bar{s}_\infty \triangleq \int_\kappa^{\lambda^+} d\mu_\lambda$$

(11)

almost surely, and

$$\lim_{(n,m) \rightarrow \infty} \bar{T}(H) = \bar{T}_\infty \triangleq \int_\kappa^{\lambda^+} \log \left(1 + \frac{\rho}{y_\lambda^+} \right) d\mu_\lambda$$

(12)

almost surely, where $d\mu_\lambda$ is defined in (7) and the constant $\kappa \in [\lambda^-, \lambda^+]$ is chosen to maximize $\bar{T}_\infty$.

Proof: Note that

$$\bar{s}(H) = \frac{1}{m} \sum_{k=1}^m g(\lambda_k)$$

is a spectral linear statistic defined through the indicator function $g(\lambda) = \chi(\lambda \geq \kappa)$. Although $g(\lambda)$ is not Lipschitz, we can approximate with a family of Lipschitz functions \{(g_k)\}, $g_k \rightarrow g$. Applying Lemma 1 to each $g_k(\lambda)$, $k = 1, 2, \ldots$, will produce (11). The proof of (12) is similar.

The above shows that, for a given SNR requirement, the optimal normalized number of on-beams $\bar{s}$ converges to a constant independent of the specific channel realization. This is reminiscent of channel hardening [33], i.e., the characteristics of a MIMO channel become deterministic as the number of transmit and receive antennas approach infinity. It is this fact which motivates a power on/off strategy with a constant number of on-beams, discussed in detail in Section III-C.

To compute or estimate the optimal threshold $\kappa$, a numerical search might be used. Starting with a grid of values of $\kappa$, compute (11) and (12), then refine the grid and recompute until the optimal $\kappa$ is well approximated. However, $T_\infty$ is not a concave function of $\kappa$ and this simple type of search can encounter obvious problems. Fortunately, we have a criterion for the optimality of $\kappa$.

Our approach follows that in [34]. We employ the bijective map from $[0, \pi]$ to $[\lambda^-, \lambda^+]$ which, with a slight abuse of notation, is defined by

$$\lambda(t) = \frac{1}{y} \left(1 + y - 2\sqrt{y \cos(t)}\right)$$

(13)

where $y \triangleq \frac{1}{\tau}$ and $t \in [0, \pi]$. It can be verified that

$$\frac{d}{dt}\lambda(t) = f_T(t) \triangleq \frac{1}{\pi} \left[1 - \frac{1 - \cos(2t)}{1 + \cos(\theta)}\right], \quad \text{if } y < 1, \quad \text{if } y = 1,$$

(14)

For any $\kappa \in [\lambda^-, \lambda^+]$, define $a = \kappa^{-1}(\kappa)$. Then

$$\bar{s}_\infty(a) = \int_\kappa^\pi f_T(t) \, dt$$

(15)

and

$$\bar{T}_\infty(a) = \int_\kappa^\pi \log \left(1 + \frac{\rho}{y_\lambda^+} (1 + y - 2\sqrt{y \cos(t)})\right) \times f_T(t) \, dt$$

(16)

It is clear that both $\bar{s}_\infty(a)$ and $\bar{T}_\infty(a)$ are differentiable functions of $a$. Denote the derivative of $\bar{T}_\infty(a)$ by $\bar{T}_\infty'(a)$. Then

$$\bar{T}_\infty'(a) = f_T(a) \left[1 - \log(1 + z(a)) - \int_a^\pi \frac{1}{1 + z(t)} \times f_T(t) \, dt\right]$$

(17)

where

$$z(t) = \frac{\rho}{y_\lambda^+} (1 + y - 2\sqrt{y \cos(t)}).$$

(18)

While $a \mapsto \bar{T}_\infty(a)$ is still not concave, we do have the following theorem.

Theorem 3:

1) $\bar{T}_\infty(a)$ has at most one root in the domain of $(0, \pi)$. If such a root exists, then $\bar{T}_\infty$ is maximized at this root.

2) If $\bar{T}_\infty(a)$ does not have any roots in $(0, \pi)$, then $\bar{T}_\infty(a) < 0$ for all $a \in (0, \pi)$ and $\bar{T}_\infty$ is maximized at $a = 0$.

Proof: See Appendix C.

Given the above, the numerical search for the optimal $\kappa$ can be simplified. Instead of building a grid of values of $\kappa$, we find $\kappa$ iteratively using the bijective mapping (13). From the proof of Theorem 3, $\bar{T}_\infty(\pi) \rightarrow -\infty$. We start with $a_1 = \pi / 2$: if $\bar{T}_\infty'(a_1) > 0$ then the optimal $a \in (a_1, \pi)$; otherwise the
optimal \( a \in (0, a_1) \). In this way, the search space is halved after each iteration. We continue the process until it stabilizes, and then we recover the optimal \( \kappa \) through (13). This iterative method not only reduces the computational cost, but also helps control the approximation error. Denote the optimal \( a \) and \( \kappa \) by \( a_{\infty} \) and \( \kappa_{\infty} \) respectively.

Corollary 1: The \( a_{\infty} \) which maximizes \( \tilde{s}_{\infty} \) is a decreasing function of \( \rho \), while the \( \tilde{s}_{\infty} \) which maximizes \( \tilde{f}_{\infty} \) is an increasing function of \( \rho \).

Proof: See Appendix D. \( \square \)

That the optimal \( s \) is increasing with \( \rho \) was partially observed for transmit antenna selection, a particular primitive case of beamforming [13]. There, at most one beam is be turned on when \( \rho \) is extremely small and \( m \) beams are turned on when \( \rho \) is sufficient large. The result of Corollary 1 is much more general as it covers all SNRs.

Note of course that to apply the asymptotic results thus far, to either search for \( \kappa_{\infty} \) or compute \( \tilde{f}_{\infty} \), the definite integrals (15)–(17) need be computed. While this may be done numerically, we derive (asymptotic) formulas for these expressions that lend themselves to real-time computation. Our method is once again similar to that in [34], though also see [35].

Asymptotic Formulas for Power On/Off Strategy: Toward computing the definite integrals (15)–(17), define

\[
\begin{align*}
\lambda &\triangleq \sqrt{y} \\
n &\triangleq \frac{\tilde{s}_{\infty} y}{\rho} \\
w &\triangleq \frac{1}{2} \left( 1 + y + \alpha + \sqrt{(1+y+\alpha)^2 - 4y} \right) \\
u &\triangleq \frac{1}{2} \left( 1 + y + \alpha - \sqrt{(1+y+\alpha)^2 - 4y} \right) \\
\theta_r &\triangleq \tan^{-1} \left( \frac{r \sin(a)}{1 - r \cos(a)} \right), \text{ if } r \cos(a) \neq 1 \text{ and} \\
\theta_u &\triangleq \tan^{-1} \left( \frac{u \sin(a)}{1 - u \cos(a)} \right), \text{ if } u \cos(a) \neq 1.
\end{align*}
\]

Also introduce three special functions:

\[
\begin{align*}
\text{Li}_2(x) &\triangleq \sum_{n=1}^{\infty} \frac{x^n}{n^2} \text{ if } |x| \leq 1 \\
\text{Sr}_1(u,r,t) &\triangleq \sum_{l=1}^{\infty} \frac{rl e^{lt}}{l} \left( \frac{\sqrt{1}}{k} \sum_{k=1}^{\infty} \frac{r^{2k} (\frac{u}{r})^k}{k} \right) \tag{20}
\end{align*}
\]

and

\[
\text{Sr}_2(r,t) \triangleq \sum_{l=1}^{\infty} \frac{rl e^{lt}}{l} \left( \frac{1}{r^{2}} \sum_{k=1}^{\infty} \frac{r^{2k}}{k} \right). \tag{21}
\]

(Li_2(x) is the usual Dilogarithm [36].)

We then have the following:

\[
\tilde{s}_{\infty} = \begin{cases} 
\frac{1}{\pi} \left\{ \frac{\pi}{\sqrt{1 - \frac{1}{r^2} \sin(a)}} + \frac{1}{\sqrt{1 - \frac{1}{r^2}}} \theta_r \right\}, & \text{if } y < 1 \\
\frac{1}{\pi} \left\{ \frac{\pi}{\sqrt{1 - \frac{1}{r^2} \sin(a)}} \right\}, & \text{if } y = 1 
\end{cases} \tag{22}
\]

\[
\tilde{f}_{\infty} = \begin{cases} 
\frac{\log(u) - \log(a)}{\log(\frac{u}{a})} \tilde{s}_{\infty} + J_0 + J_1 + J_2, & \text{if } y < 1 \\
\frac{\log(u) - \log(a)}{\log(\frac{u}{a})} \tilde{s}_{\infty} + J_0 + J_1, & \text{if } y = 1 \tag{23}
\end{cases}
\]

where

\[
J_0 = \frac{1}{\pi r} \left\{ \sin(a) [1 - \log \left( 1 + u^2 - 2u \cos(a) \right)] - u(\pi - a) + \frac{1}{u} \theta_u \right\}
\]

and

\[
J_1 = \frac{1+y}{2 \pi y} \left\{ \text{Li}_2 \left( \frac{ue^{-iu}}{a} \right) - \text{Li}_2 \left( \frac{ue^{-iu}}{a} \right) \right\},
\]

and

\[
J_2 = \frac{1-y}{2 \pi y} \left[ - 2 \log(1 - ur)(\pi - a - \theta_r) + i \text{Sr}_1(u,r,a) - i \text{Sr}_1(u,r,-a) \right];
\]

\[
\tilde{f}'_{\infty}(a) = \frac{J_3}{\pi} \left[ 1 - \log \left( 1 + \frac{\rho}{\tilde{s}_{\infty} y} \right) \times \left( 1 + y^2 - 2y \cos(a) \right) - \frac{y}{\rho} \right] \tag{24}
\]

where

\[
J_3 = \begin{cases} 
\frac{1 - \cos(2a)}{1 + \frac{1}{r^2} - 2r \cos(a)} & \text{if } y < 1 \\
1 + \cos(a), & \text{if } y = 1
\end{cases}
\]

and

\[
\tilde{f}' = \begin{cases} 
\frac{1}{\pi \log(1 - ur)} \left[ \frac{\pi}{\sqrt{1 - \frac{1}{r^2}} \sin(a)} + \frac{1}{\sqrt{1 - \frac{1}{r^2}}} \theta_r \right], & \text{if } y < 1 \\
\frac{1}{\pi \log(1 - ur)} \left( \frac{(1+i)\theta_r}{(1+i)\theta_r} \right), & \text{if } y = 1
\end{cases}
\]

The derivations themselves are somewhat technical and uninformative. We omit them here, but the details may be found in the technical report [37].

Despite their cumbersome form, the above extend various results in [34], [35], [38], and [39], which are valid only when \( a = 0 \). (Setting \( a = 0 \) in our formulas produces significant simplifications and recovers those earlier forms, as it must.) Most noteworthy is that the feedback gain is more significant at low SNRs and the threshold \( a \) increases as SNR decreases. Thus, our results are more germane to a finite rate feedback analysis. Further, the infinite series (19)–(21) need only be considered when \( y < 1 \), in which case they converge faster than the corresponding geometric series. Hence, one need only retain a small number of terms for a sharp estimate (with a quantifiable error), making the presented expressions suitable for real-time implementation.

Asymptotic Formulas for CSITR Case: As a byproduct, we also present asymptotic formulas to compute the capacity achieved by power water-filling and the corresponding Lagrange multiplier. Assume the same asymptotic regime as before.

We have

\[
\tilde{C}_{\infty} \triangleq \lim_{(m,\nu) \to \infty} \frac{1}{m} C = \int_{\max(\lambda,1/\nu)}^{\lambda^*} \log(\nu) d\nu
\]
where \( \nu \) is the Lagrange multiplier satisfying the power constraint

\[
\rho = \int_{\max(\lambda-1/2\nu)}^{\lambda^+} \left( \nu - \frac{1}{\lambda} \right) d\mu_\lambda.
\]

Apply the same change of variables used in (13). It is clear that

\[
\bar{C}_\infty = \int_a^\pi \log \left( \frac{\nu}{y} (1 + y - 2\sqrt{y} \cos(t)) \right) f_T(t) \, dt,
\]

where

\[
a = \begin{cases} \cos^{-1} \left( \frac{1 + \nu - \sqrt{\nu}}{2\sqrt{\nu}} \right), & \text{if } \lambda^- \leq \frac{1}{\nu} \leq \lambda^+ \\ 0, & \text{if } \frac{1}{\nu} < \lambda^-
\end{cases}
\]

and

\[
\rho = \int_a^\pi \left( \nu - \frac{y}{1 + y - 2\sqrt{y} \cos(t)} \right) f_T(t) \, dt.
\]

Then we obtain the following.

- The Lagrange multiplier \( \nu \) is given by

\[
\nu = \bar{C}_\infty - J_4
\]

where \( \bar{C}_\infty \) is defined in (22), we have the equation shown at the bottom of the page and

\[
a = \begin{cases} \cos^{-1} \left( \frac{1 + \nu - \sqrt{\nu}}{2\sqrt{\nu}} \right), & \text{if } \lambda^- \leq \frac{1}{\nu} \leq \lambda^+ \\ 0, & \text{if } \frac{1}{\nu} < \lambda^-
\end{cases}
\]

- \( \bar{C}_\infty = \begin{cases} \log \left( \frac{\nu}{y} \right) \bar{C}_\infty + J_5 + J_6 + J_7, & \text{if } y < 1 \\ \log \left( \frac{\nu}{y} \right) \bar{C}_\infty + J_6 + J_0, & \text{if } y = 1
\end{cases}
\]

where \( \bar{C} \) and \( a \) are evaluated by (22) and (26), respectively

\[
J_5 = \frac{1}{\pi r} \left\{ \sin(a) \left[ 1 - \log(1 + r^2 - 2r \cos(a)) \right] - r(\pi - a) - \left( 1 - r - r_a \right) \theta_r \right\}
\]

\[
J_6 = \frac{1 + r^2}{2\pi r^2} \left[ L_2(r e^{-i\alpha}) - L_2(r e^{i\alpha}) \right]
\]

and

\[
J_7 = -\frac{1 - r^2}{2\pi r^2} \times \left\{ \frac{i}{2} \left[ \log^2 \left( 1 - r e^{-i\alpha} \right) - \log^2 \left( 1 - r e^{i\alpha} \right) \right] 
+ 2 \log \left( 1 - r^2 \right) \left( \pi - a - \theta_r \right) 
+ i \left[ S_2 \left( r, -a \right) - S_2 \left( r, a \right) \right] \right\}.
\]

Once more, the proofs are contained in the technical report [37] and are omitted here.

### C. MIMO Systems With Finite Many Antennas

It is by now well understood that asymptotic results are often sufficiently accurate to apply in MIMO systems with only a few antennas [33]–[35], [40]. The same is true of the results of Section III-B. Theorem 2 proves that the optimal normalized number of on-beams \( \bar{s} \) converges to a constant almost surely (in our limiting regime). We now demonstrate that fixing \( \bar{s} \) is near optimal even when the numbers of antennas are small.

Before proceeding, note the fundamental difference between the asymptotic and finite cases. In the asymptotic case, \( \bar{s} \) may live throughout \([0,1]\). When the numbers of antennas is finite, \( \bar{s} \) can only take finitely many discrete values, i.e., \( \bar{s} \in \{ \frac{1}{m}, \frac{2}{m}, \ldots, 1 \} \). So, to apply the asymptotic results to the finite case, we must “quantize” in the self-evident manner.

1. For a given MIMO system and a given SNR \( \rho \), evaluate the asymptotically optimal \( a_\infty \) and \( \bar{s}_\infty \) as spelled out in Section III-B.

2. If \( \bar{s}_\infty < \frac{1}{m} \), go to \( 3 \). Otherwise, choose the integer \( s \) maximizing \( T \) from the two integers adjacent to \( m\bar{s}_\infty \). Specifically, let \( s_1 = \frac{1}{m} \lfloor m\bar{s}_\infty \rfloor \) and \( s_2 = \frac{1}{m} \lceil m\bar{s}_\infty \rceil \). Compare the corresponding performances \( T_\infty \) and select the \( \bar{s} \in \{ s_1, s_2 \} \) maximizing \( T_\infty \). The on-power is given by \( P_{on} = \rho / (m\bar{s}) \).

3. If \( \bar{s}_\infty < \frac{1}{m} \), then we turn on/off the strongest eigenchannel according to the threshold test

\[
\lambda_1 \geq \frac{\kappa}{m\bar{s}_\infty}
\]

where \( \kappa = \frac{1}{\nu} (1 + y - 2\sqrt{y} \cos(a_\infty)) \) and keep all other eigenchannels off. The on-power can be approximated by \( P_{on} \approx \frac{\rho}{m\bar{s}} \).

The power on/off strategy outlined just now is a power on/off strategy with a constant number of on-beams. When SNR \( \rho \) is not very low so that \( \bar{s}_\infty \geq \frac{1}{m} \), the number of on-beams is a constant independent of the specific channel realization \( \mathbf{H} \). The only exception occurs when \( \rho \) is so low that \( \bar{s}_\infty < \frac{1}{m} \); the strongest beam is turned on if \( \lambda_1 \geq \kappa \); otherwise no beams are on. In this case, the exact on-power can be calculated by simulation or adapted on-line; we use an asymptotic approximation for simplicity.

Empirical simulations show that the proposed strategy is near-optimal for MIMO systems with finitely many antennas. Simulated information rate versus SNR is presented in Fig. 2 while Fig. 3(a) shows simulated information rate versus \( E_b/N_0 \). MIMO systems with 4 \( \times \) 2, 4 \( \times \) 3, and 4 \( \times \) 4 antennas are considered. The solid line and the dashed line are the simulated
Fig. 2. Information rate versus SNR for perfect beamforming.

Fig. 3. Information rate versus $E_b/N_0$ for perfect beamforming.

IV. POWER ON/OFF STRATEGY WITH A FINITE SIZE BEAMFORMING CODEBOOK

In the previous section, we have assumed perfect beamforming in order to decouple the effect of power on/off and beamforming. With a finite feedback rate, it is impossible to always choose a beamforming matrix to perfectly match the right singular vectors of the channel state matrix. We will now characterize the performance loss due to this imperfect beamforming.

We focus on the proposed power on/off strategy with a constant number of on-beams. Let $s$ be the number of on-beams. In the proposed strategy, the receiver needs to feedback information regarding the first $s$ right singular vectors of the matrix $H$. Let $R_n$ be the feedback rate available. A beamforming codebook of the form

$$B = \left\{ Q_i \in \mathbb{C}^{L \times s} : \text{Tr}(Q_i) = 1, 1 \leq i \leq 2^{R_n} \right\}$$

(linearly with the system dimension $m$ for a given $\rho$, i.e., $T = m \left( T + o_m(1) \right)$, where $o_m(1) \rightarrow 0$ as $m \rightarrow \infty$. The constant $T$ can also be well approximated, for all SNR, by the same asymptotic analysis.)
should be declared to both the transmitter and the receiver. The only exception is the case when SNR is so low that $\frac{1}{m}$ (see Section III-C for details). In this case, the feedback is either the index of a beamforming matrix or an extra index to indicate power off. The cardinality of the corresponding beamforming codebook decreases by one, i.e.

$$B = \{ Q_i \in C^{L_T \times 1} : Q_i^\dagger Q_i = 1, 1 \leq i \leq 2^{R_B} - 1 \}. \quad (29)$$

We shall study the design problem in (5) associated with a single rank beamforming codebook. That is, a codebook containing beamforming matrices all of the same rank.

To solve the corresponding optimization problem and make the performance analysis tractable, an asymptotic optimal feedback function is introduced and beamforming codebook design is discussed in Section IV-A. The achievable information rate is characterized in Section IV-B.

A. Feedback Function and Beamforming Codebook Design

Given a single rank beamforming codebook, the optimal feedback function $\varphi^*(\cdot)$ is clearly given by

$$\varphi^*(H) = \arg \max_{1 \leq i \leq [s]} \log \left( I_{L_R} + P_{en} H Q_i H^\dagger \right).$$

Yet again, the corresponding performance analysis is too complicated to be tractable. We therefore introduce a suboptimal feedback function.

Consider the singular value decomposition $H = U A V^\dagger$. Let $V_s$ denote the $L_T \times s$ matrix composed of the right singular vectors corresponding to the largest $s$ singular values. Recall the definitions of the Stiefel and Grassmann manifolds from Section II-C. Both $V_s, Q \in S_{L_T,s}(\mathbb{C})$ span $s-$ dimensional planes in $C^{L_T}$, denoted by $\mathcal{P}(V_s)$ and $\mathcal{P}(Q)$, respectively. Our feedback function $\hat{\varphi}(\cdot)$ is based on the chordal distance defined on the $G_{L_T,s}(\mathbb{C})$:

$$\hat{\varphi}(H) \triangleq \arg \min_{1 \leq i \leq [s]} \min_{Q_s \in B_{[s]}} d_c(\mathcal{P}(Q_i), \mathcal{P}(V_s)).$$

$$= \arg \max_{1 \leq i \leq [s]} \log \left( \left| V_s^\dagger Q_i V_s \right|^2 \right). \quad (30)$$

The feedback function (30) is asymptotically optimal as $R_B \rightarrow \infty$. Indeed, since $G_{L_T,s}(\mathbb{C})$ is compact, for any $\epsilon > 0$ and $R_B$ is sufficiently large, there exits a codebook $B \subset S_{L_T,s}(\mathbb{C})$ such that $|B| \leq 2^{R_B}$ and $\mathcal{P}(Q) \in G_{L_T,s}(\mathbb{C})$. Form an $\epsilon$-net in $G_{L_T,s}(\mathbb{C})$ with respect to the chordal distance. Let $R_B \rightarrow \infty$ and $B_{[s]}$ be a family of codebooks such that

$$\lim_{R_B \rightarrow \infty} \sup_{Q_s \in B_{[s]}} \min_{Q_{1\leq i \leq [s]}} d_c(\mathcal{P}(Q_i), \mathcal{P}(V_s)) = 0.$$

Following the same argument in the proof of Lemma 2, it can be shown that the information rate achieved by the sub-

optimal feedback function $\hat{\varphi}(\cdot)$ approaches that of perfect beamforming. The asymptotic optimality of $\hat{\varphi}(\cdot)$ is therefore established.

We next record an important property of the feedback function $\hat{\varphi}(\cdot)$.

Theorem 4: Let $B$ be a single rank beamforming codebook with rank $1 \leq s \leq L_T$. Let $V_s \in S_{L_T,s}(\mathbb{C})$ be a random matrix randomly drawn from the isotropic distribution. Let

$$\hat{\varphi}(V_s) = \arg \min_{1 \leq i \leq [s]} d_c(\mathcal{P}(Q_i), \mathcal{P}(V_s)).$$

Then

$$E V_s \left[ V_s^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger V_s \right] = \mu I$$

where

$$\mu \triangleq 1 - \frac{1}{s} E V_s \left[ d_c^2(\mathcal{P}(Q_{\hat{\varphi}(V_s)}), \mathcal{P}(V_s)) \right] \geq 0. \quad (31)$$

To prove Theorem 4, we need the following lemma.

Lemma 3: Let $A \in \mathbb{C}^{n \times k}$ be a Hermitian matrix. If $A = U A U^\dagger$ for all $k \times k$ unitary matrix $U$, then $A = \mu I$ for some constant $\mu \in \mathbb{R}$.

Proof: For any Hermitian $A$, there exists a $k \times k$ unitary $U$ such that $U A U^\dagger = A$ where $A$ is diagonal and with real diagonal elements. But $U A U^\dagger = A$, then $A$ is diagonal and real. Choosing a permutation matrix for $U$ shows that the diagonal elements are identical.

Proof (Proof of Theorem 4): For any given $s \times s$ unitary matrix $U$, $\varphi(V_s, U) = \hat{\varphi}(V_s)$ because $\mathcal{P}(V_s) = \mathcal{P}(V_s U)$. Since $V_s$ is isotropically distributed, so is $V_s U[31]$. Then

$$U^\dagger E V_s \left[ V_s^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger V_s \right] U$$

(a) $\Rightarrow E V_s \left[ (V_s U)^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger (V_s U) \right]$ (b) $\Rightarrow E V_s \left[ (V_s U)^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger (V_s U) \right]$ (c) $\Rightarrow E V_s \left[ V_s^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger V_s \right]$ where (a) follows from the fact that $\hat{\varphi}(V_s U) = \hat{\varphi}(V_s)$, (b) follows as $V_s U$ and $V_s$ have the same distribution, and (c) is a variable change from $V_s U$ to $V_s$. Therefore

$$E V_s \left[ V_s^\dagger Q_{\hat{\varphi}(V_s)} Q_{\hat{\varphi}(V_s)}^\dagger V_s \right] = \mu I$$

for some constant $\mu$ according to Lemma 3. It is elementary to show that

$$\mu = 1 - \frac{1}{s} E V_s \left[ d_c^2(\mathcal{P}(Q_{\hat{\varphi}(V_s)}), \mathcal{P}(V_s)) \right] \geq 0. \quad \square$$

The constant $\mu$ introduced in Theorem 4 is related to the distortion (squared chordal distance) of quantization on the Grassmann manifold. Particularly, we are interested in the maximum
Achievable given a codebook size. This problem is studied in our companion paper [25] and we cite the relevant results in the next theorem.

**Theorem 5.** Let \( B \) be a single rank beamforming codebook with \( 1 \leq s \leq L_T \). Let \( V_s \in S_{L_T,s}(\mathbb{C}) \) be isotropically distributed. Consider the average quantization performance associated with the codebook \( B \) described by

\[
D_c(B) \triangleq E_{V_s} \left[ \min_{Q \in B} \mathcal{D}_R^2(\mathcal{P}(Q), \mathcal{P}(V_s)) \right].
\]

For a given \( K \in \mathbb{Z}^+ \), define the minimum achievable performance by

\[
D_c^*(K) \triangleq \inf_{B} D_c(B).
\]

Then, Assume that \( K \) is large. \( D_c^*(K) \) is bounded by

\[
\frac{t}{t+1} \eta^{-1/2} \log_2 K/t \left( 1 + o_{L_T}(1) \right) \leq D_c^*(K) \leq \frac{1}{t} \eta^{-1/2} \log_2 K \left( 1 + o_{L_T}(1) \right)
\]

where \( t = s(L_T - s) \) is the number of the real dimensions of \( G_{L_T,s}(\mathbb{C}) \).

\[
\eta = \begin{cases} \frac{1}{t} \prod_{j=1}^s \left( \frac{L_T - s_j}{s_j - 1} \right), & \text{if } 1 \leq s \leq \frac{L_T}{2} \\ \frac{1}{t} \prod_{j=1}^{L_T-s} \left( \frac{L_T - s_j}{s_j - 1} \right), & \text{if } \frac{L_T}{2} < s \leq L_T \end{cases}
\]

and \( o_{L_T}(1) \rightarrow 0 \) when \( s \) is fixed and \( K, L_T \rightarrow \infty \) with \( \frac{1}{t} \log_2 K \rightarrow c \in \mathbb{R}^+ \).

The tightness of the bounds in (32) are demonstrated by their asymptotic identity as \( s \) is fixed and \( K, L_T \rightarrow \infty \) with \( \frac{1}{t} \log_2 K \rightarrow c \in \mathbb{R}^+ \). In practice, we ignore the multiplicative \((1 + o_{L_T}(1))\) errors and use just the leading order term to approximate the function \( D_c^*(K) \). This approximation is good even when \( K \) is relatively small. For example, when \( L_T = 4 \) and \( s = 2 \), both bounds approximate \( D_c^*(K) \) well for all \( K \geq 10^{25} \).

Applying Theorem 5 to Theorem 4, the maximum \( \mu \) achievable, say \( \mu_{\text{sup}} \), can be upper and lower bounded by

\[
1 - \frac{1}{s(t+1)} \eta^{-1/2} \log_2 R_b/t \leq \mu_{\text{sup}} \leq 1 - \frac{t}{s(t+1)} \eta^{-1/2} \log_2 R_b/t
\]

where the symbol \( \ll \) indicates we have dropped the \((1 + o_{L_T}(1))\) corrections. Here, we have that \( \log_2 |B| = R_b \), when \( s c_{\infty} \geq \frac{m}{2} \), then \( \log_2(2^{R_b} - 1) \approx R_b \), where \( c_{\infty} < \frac{m}{2} \) and \( R_b \in \mathbb{Z}^+ \) is not small. Based on these bounds, we are able to approximate the achievable information rate, which is the main topic of the next section.

The problem of how to design a codebook to minimize the distortion is still not completely solved. The obtained bounds in (32) suggest a randomly generated codebook might be suitable. We henceforth consider codebooks with elements drawn independently from the isotropic distribution. In fact, the upper bound in (32) is obtained by computing the average distortion of such random codebooks, and the asymptotic identity of the lower and upper bounds implies that random codes are asymptotically optimal in average. The authors also proved a stronger result that random codes are asymptotically optimal in probability [25]. When the codebook size is finite, a randomly generated codebook may not be optimal, but typically performs well enough. There are several proposed techniques for beamforming codebook design in the literature [15], [19]. These usually attempt to maximize the minimum chordal distance between the planes spanned by any pair of beamsforming matrices (the max-min criterion). The resulting codebooks perform on average slightly better than a randomly generated codebook when the codebook size is small, but become worse when the codebook is large. Still, the max-min criterion has been widely adopted in practice and will be used in our simulations.

**B. Achievable Information Rate**

The information rate of power on/off strategy with finite rate feedback is characterized by combining the asymptotic results in Section III and the results in Section IV-A on channel quantization.

First we derive a lower bound. For a channel state realization \( \mathbf{H} \), let \( \lambda_i \) be the \( i \)-th largest eigenvalue of \( \mathbf{H}^{\dagger} \mathbf{H} \) and \( \mathbf{v}_i \) be the corresponding eigenvector. Let \( \mathbf{V}_s = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s] \) and \( \mathbf{A}_s = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_s] \). Then

\[
\mathbf{V} \mathbf{A} \mathbf{V}^{\dagger} \geq \mathbf{V} \begin{bmatrix} \mathbf{A}_s & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^{\dagger} = \mathbf{V} \mathbf{A}_s \mathbf{V}^{\dagger}_s \quad (35)
\]

where the matrix relationship \( \mathbf{A} \geq \mathbf{B} \) means that \( \mathbf{A} - \mathbf{B} \) is nonnegative definite. Let \( \mathbf{Q}_{\phi}(\mathbf{H}) \) be the feedback beamforming matrix given by the feedback function (30). Left and right multiply both sides of (35) by \( \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \) and \( \mathbf{Q}_{\phi}(\mathbf{H}) \), respectively, and then add both sides by \( \mathbf{I}_s \). We have

\[
\mathbf{I}_s + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A} \mathbf{V}^{\dagger} \mathbf{Q}_{\phi}(\mathbf{H}) \geq \mathbf{I}_s + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A}_s \mathbf{V}^{\dagger}_s \mathbf{Q}_{\phi}(\mathbf{H}) \quad (36)
\]

Moreover, the matrices on both sides of the above inequality are positive definite. Because \( \mathbf{A} \geq \mathbf{B} \geq 0 \) implies \( |\mathbf{A}| \geq |\mathbf{B}| \), we have

\[
\log \left| \mathbf{I} + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A} \mathbf{V}^{\dagger} \mathbf{Q}_{\phi}(\mathbf{H}) \right| \\

greater \log \left| \mathbf{I} + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A}_s \mathbf{V}^{\dagger}_s \mathbf{Q}_{\phi}(\mathbf{H}) \right|
\]

Therefore, the information rate is lower bounded by

\[
\mathcal{I} = \frac{1}{m} \mathbb{E}_\mathbf{H} \left[ \log \left| \mathbf{I}_{L_{\text{th}}} + P_{\text{on}} \mathbf{H} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{Q}_{\phi}(\mathbf{H}) \mathbf{H}^{\dagger} \right| \right] \\
= \frac{1}{m} \mathbb{E}_\mathbf{H} \left[ \log \left| \mathbf{I}_s + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{H}^{\dagger} \mathbf{H} \mathbf{Q}_{\phi}(\mathbf{H}) \right| \right] \\
= \frac{1}{m} \mathbb{E}_\mathbf{H} \left[ \log \left| \mathbf{I}_s + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A} \mathbf{V}^{\dagger} \mathbf{Q}_{\phi}(\mathbf{H}) \right| \right] \\
\geq \frac{1}{m} \mathbb{E}_\mathbf{H} \left[ \log \left| \mathbf{I}_s + P_{\text{on}} \mathbf{Q}_{\phi}(\mathbf{H})^{\dagger} \mathbf{V} \mathbf{A}_s \mathbf{V}^{\dagger}_s \mathbf{Q}_{\phi}(\mathbf{H}) \right| \right]
\]

where equality holds if we have perfect beamforming \( \mathbf{Q}_{\phi}(\mathbf{H}) \mathbf{V}_s = \mathbf{I}_s \).

Based on this lower bound, an approximation to the information rate can be obtained. Since entries of \( \mathbf{H} \) are i.i.d. \( \mathcal{C}\mathcal{N}(0, 1) \),
$V_s \in S_{L_T,s}(\mathbb{C})$ is isotropically distributed and independent with $A_s$ \cite{30}, \cite{31}. By the lower bound in \eqref{eq:36}, we have

$$
\frac{1}{m} F_{\mathbf{H}} \left[ \log \left[ I_s + P_{\text{on}}Q_{\phi}(\mathbf{H}) V_s A_s V_s^\dagger Q_{\phi}(\mathbf{H}) \right] \right]
\geq \frac{1}{m} E_{A_s} \left[ E_{V_s} \left[ \log \left[ I_s + P_{\text{on}} V_s^\dagger Q_{\phi}(\mathbf{H}) Q_{\phi}(\mathbf{H}) V_s A_s \right] \right] \right]
= \frac{1}{m} E_{A_s} \left[ \log \left[ I_s + P_{\text{on}} E_{V_s} \left[ V_s^\dagger Q_{\phi}(\mathbf{H}) Q_{\phi}(\mathbf{H}) V_s \right] A_s \right] \right]
\geq \frac{1}{m} E_{\mathbf{H}} \left[ \log \left[ I + \mu P_{\text{on}} A_s \right] \right]
$$

\begin{equation}
(37)
\end{equation}

where (a) holds because $[I + AB] = [I + BA]$ and $V_s$ is independent of $A_s$, (b) follows from the concavity of $\log \cdot$ function \cite[Prob. 2 on p. 237]{41}, and (c) is just the definition of $\mu$ in \eqref{eq:31}.

Although the derived \eqref{eq:37} is only an approximation of the normalized information rate, the equality in \eqref{eq:37} holds when beamforming is perfect ($\mu = 1$). As we shall show later, it is fairly accurate whenever the feedback rate is not small.

Remark 1: We refer to the constant $\mu$ as the power efficiency factor. It is the single scalar that captures the effect of imperfect beamforming as the on-power $P_{\text{on}}$ is effectively decreased to $\mu P_{\text{on}}$ because of imperfect beamforming. The power loss is proportional to

$$
2^{-B_{\text{th}}/s(L_T-s)}.
$$

This means that the power loss decreases exponentially with the feedback rate $B_{\text{th}}$. Our simulations demonstrate that for systems with small numbers of antennas, a few bits of feedback introduce a significant gain while the gain obtained by further increasing the feedback rate is marginal.

One advantage of our proposed strategy, compared with the schemes in which the channel matrix $\mathbf{H}$ is quantized directly, is dimension reduction. The number of real dimensions of $P(V_s) \in G_{L_T,s}(\mathbb{C})$ is $2s(L_T-s)$ while that of $\mathbf{H} \in \mathbb{C}^{L_T \times L_R}$ is $2L_T L_R$. Our proposed strategy only quantizes the hyperplane generated by $V_s$, and therefore significantly reduces the dimensions of the quantization space and save the feedback resource.

To estimate the achievable information rate, we use \eqref{eq:34} to estimate the maximum achievable $I_{\text{approx}}$, substitute it into the approximation \eqref{eq:37}, and then apply our asymptotic results to calculate \eqref{eq:37}. Although this process involves several approximations, the obtained estimate works well in practice. In fact, for a $4 \times 2$ MIMO system with feedback rate $B_{\text{th}} = 4$ bits/channel use, our theoretical estimate is very close to that from Monte Carlo simulations.

Fig. 4 gives the simulation results for a $4 \times 2$ MIMO system. To simulate the effect of imperfect beamforming, single rank beamforming codebooks are constructed: we start with a randomly generated codebook, iteratively move the codewords to maximize their minimum distance (the max-min criterion). The performance curves are plotted as functions of $R_{\text{th}}/m^2$. The simulated information rate (circles) is compared to the information rate calculated by using bounds on the minimum achievable quantization distortion \eqref{eq:32}. The simulation results show that the information rate based on the bounds \eqref{eq:32} matches the actual performance almost perfectly.

V. PERFORMANCE COMPARISON

While we have shown that fixing the number of on-beams is near optimal for perfect beamforming in Section III, this section will show that it is actually near optimal for imperfect beamforming as well.

To set the benchmark, we need a power on/off strategy in which the number of on-beams is allowed to vary with the channel realization. Such a strategy requires a beamforming codebook containing beamforming matrices of different ranks. Refer to such a codebook as a multirank beamforming codebook. It is a union of several single rank subcodes

$$
B = \bigcup_{s=0}^{L_T} B_s
$$

where $B_s \subset S_{L_T,s}(\mathbb{C})$.

The corresponding system design problem is the same as the one stated in \eqref{eq:5}. The fact that the codebook $B$ may contain beamforming matrices with different ranks makes the problem more complicated. To simplify the problem, let us fix the codebook $B$ and the on-power $P_{\text{on}}$, and focus on the design of the feedback function. This optimal feedback function is obtained in the next theorem. Similar to before, the intuition is that all the “good” beams and only the “good” beams should be turned on, we just need a new concept of “good beams.”

Theorem 6: Consider the power on/off strategy with a given nontrivial beamforming codebook $B = \bigcup_{s=0}^{L_T} B_s (3s > 0)$ such that $B_s \neq \phi$ and a given feasible $P_{\text{on}} \in \mathbb{R}^+$. For a given channel realization $\mathbf{H}$, define $T_s(\mathbf{H})$ as the largest information rate achieved by the subcode $B_s$

$$
T_s(\mathbf{H}) = \max_{Q_s \in B_s} \log \left[ I_{L_R} + P_{\text{on}} HH^\dagger Q_s H^\dagger \right]
$$

\begin{equation}
(39)
\end{equation}

where $T_s(\mathbf{H}) = 0$ if $B_s = \phi$ or $s = 0$. Then the optimal feedback function is given by

$$
\phi(\mathbf{H}) = \arg \max_{\bar{s} \in B_{\text{th}}} \log \left[ I_{L_R} + P_{\text{on}} HH^\dagger Q_s H^\dagger \right]
$$

\begin{equation}
(40)
\end{equation}
where
\[ \mathcal{S}(\mathbf{H}) \triangleq \max \left\{ s : \mathcal{H}_s(\mathbf{H}) - \mathcal{H}_t(\mathbf{H}) \geq (s-t)\kappa \forall 0 \leq t < s \right\} \]
and \( \kappa \) is the appropriate threshold such that
\[ E_{\mathbf{H}}[\mathcal{S}(\mathbf{H}) P_{\text{on}}] = \rho. \]

**Proof:** See Appendix E. \( \square \)

The following examples are direct applications of Theorem 6.

**Example 1:** Let \( \mathcal{B} = \{ \mathbf{I}_{LT}, \mathbf{Q}_0 \} \) where the symbol \( \mathbf{Q}_0 \) indicates power off. Then the optimal power on/off function is to turn on all transmit antennas if
\[ \log (\mathbf{I}_{LT} + P_{\text{on}} \mathbf{H} \mathbf{H}^H) \geq \kappa L_T \]
and turn off the transmitter if
\[ \log (\mathbf{I}_{LT} + P_{\text{on}} \mathbf{H} \mathbf{H}^H) < \kappa L_T \]
where \( \kappa \) is an appropriate chosen threshold to satisfy
\[ L_T P_{\text{on}} \Pr \left\{ \log (\mathbf{I} + P_{\text{on}} \mathbf{H} \mathbf{H}^H) \geq \kappa L_T \right\} = \rho. \]

**Example 2:** Let \( |\mathcal{B}| \rightarrow \infty \) and \( \mathcal{B} \) is constructed so that the beamforming is asymptotically perfect. It is easy to verify that the optimal feedback function given by Theorem 6 is same as the one given in Theorem 1 for perfect beamforming case.

Although the optimal feedback function is obtained, it is not clear how to construct and analyze a multirank beamforming codebook, or how to find the optimal on-power. In order to compare the performance of single-rank and multi-rank beamforming codebooks, we numerically search for the best possible solution. Specifically, denote \( K_s \) as the size of the subcode \( \mathcal{B}_s \), \( K_s \triangleq |\mathcal{B}_s| \). We try all possible combinations of \( [K_0, K_1, \ldots, K_{LT}] \) such that \( K_s \in \mathbb{Z}^+ \cup \{0\} \) and \( \sum_{s=0}^{L_T} K_s \leq \mathcal{R}^{\text{fe}} \). For each \( [K_0, K_1, \ldots, K_{LT}] \), we design subcodes \( \mathcal{B}_s \) such that \( |\mathcal{B}_s| = K_s \) for \( s = 0, 1, \ldots, L_T \) using the max-min criterion\(^3\), and employ them to form the overall codebook \( \mathcal{B} \). For every multirank codebook \( \mathcal{B} \), we apply different \( P_{\text{on}} \) and the optimal feedback function \( \bar{\varphi} (\cdot) \) in (40). We simulate the corresponding SNR \( \rho \) and information rate \( I \). For each SNR, we choose the multirank beamforming codebook to maximize the information rate.

Our empirical comparison of single-rank and multirank beamforming codebooks is presented in Fig. 5, with Fig. 5(b) focusing on the relative performance defined as the achieved information rate normalized by the capacity of a \( 4 \times 2 \) MIMO system with perfect CSIT. Simulations show that single rank beamforming codebooks (dashed lines) achieve almost the same information rate of multirank beamforming codebooks (circles). Noticeable differences in the relative performance only occur at very low SNRs, due to the two different methods used to calculate the on-power \( P_{\text{on}} \). For multirank beamforming codebooks, the \( P_{\text{on}} \) value is numerically optimized. However,

\(^3\)As discussed at the end of Section IV-A, the max-min criterion may not result in the optimal subcodes. Nevertheless, it provides us reasonably good subcodes when the code size is small.

**VI. CONCLUSION**

This paper considers the design problem for MIMO systems with limited feedback and proposes a power on/off strategy with a constant number of on-beams. The proposed strategy has low complexity and is near optimal for a large range of SNRs. The effects of power on/off and beamforming are studied through asymptotic random matrix theory and quantization results on the Grassmann manifold. Theoretical formulas are derived to approximate the achievable information rate and demonstrated to be accurate.
Theorem 7 (Modified Version of [42, Th. 3.3.4]): For given matrices $A \in \mathbb{C}^{s \times L_T}$ and $B \in \mathbb{C}^{L_T \times s}$ where $0 < s \leq L_T$, denote the ordered singular values of $A$, $B$, and $AB$ by $\lambda_1 (A) \geq \cdots \geq \lambda_s (A) \geq 0$, $\lambda_1 (B) \geq \cdots \geq \lambda_{L_T} (B) \geq 0$, and $\lambda_1 (AB) \geq \cdots \geq \lambda_s (AB) \geq 0$. Then

$$\prod_{i=1}^{s} \lambda_i (AB) \leq \prod_{i=1}^{s} \lambda_i (A) \lambda_i (B).$$

Combine (41) and (42). We then have

$$\log |I + (P_{on} \mathbb{H} \mathbb{Q}^{\dagger} \mathbb{H}^\dagger)| \leq \sum_{i=1}^{s} \log (1 + m P_{on} \lambda_i).$$

(43)

It is easy to verify that the upper bound (43) is achieved by setting $\mathbb{Q} = \mathbb{V}_s$. This completes the proof.

B. Proof of Theorem 1

Describe an arbitrary power on/off strategy with on-power $P_{on}$ by defining the events

$$\Omega_k = \{ \lambda : k^{th} \text{ eigenchannel is on} \} \quad k = 1, 2, \ldots, L_T$$

in terms of which the power constraint may be expressed as

$$\frac{P}{P_{on}} = \int_{\Omega_k} \left[ \sum_{k=1}^{L_T} \chi \lambda_k \right] d\mu (\lambda) = \sum_{k=1}^{L_T} \int_{\Omega_k} d\mu (\lambda).$$

Here, $\chi_A$ denotes the indicator of the event $A$ and we are using $\mu$ for the joint eigenvalue law at dimension $L_T$ (this is not the limiting spectral measure from (7)). The corresponding information reads

$$I = \sum_{k=1}^{L_T} \int_{\Omega_k} \log (1 + P_{on} \lambda_k) d\mu (\lambda).$$

Next consider our proposed on/off strategy in which a beam is on or off depending on whether $\lambda_k \geq \kappa$ or $\lambda_k < \kappa$, and $\kappa$ is chosen so that

$$\frac{P}{P_{on}} = \sum_{k=1}^{L_T} \int_{\Omega_k} d\mu (\lambda).$$

We can express this in a similar way to the above by defining

$$\Omega_k^* = \{ \lambda \in \mathbb{R}^{L_T} : \lambda_k \geq \kappa \}, \quad k = 1, 2, \ldots, L_T$$

and writing

$$I^* = \sum_{k=1}^{L_T} \int_{\Omega_k^*} \log (1 + P_{on} \lambda_k) d\mu (\lambda).$$

for the power constraint and

$$I^* \geq I.$$ for the corresponding information rate.

We must show that $I^* \geq I$. Start by noting that

$$I^* - I = \sum_{k=1}^{L_T} \int_{\Omega_k^*} \log (1 + P_{on} \lambda_k) d\mu (\lambda) - \sum_{k=1}^{L_T} \int_{\Omega_k} \log (1 + P_{on} \lambda_k) d\mu (\lambda)$$

$$= \sum_{k=1}^{L_T} \int_{\Omega_k^*} \log (1 + P_{on} \lambda_k) d\mu (\lambda)$$

$$- \sum_{k=1}^{L_T} \int_{\Omega_k - \Omega_k^*} \log (1 + P_{on} \lambda_k) d\mu (\lambda).$$

By the definition of $\Omega_k^*$

$$\log (1 + P_{on} \lambda_k) \geq \log (1 + P_{on} \lambda_k) \forall \lambda \in \Omega_k^* - \Omega_k$$

and

$$\log (1 + P_{on} \lambda_k) \leq \log (1 + P_{on} \lambda_k) \forall \lambda \in \Omega_k - \Omega_k^*.$$

Thus

$$I^* - I \geq \log (1 + P_{on} \kappa) \sum_{k=1}^{L_T} \left( \int_{\Omega_k^* - \Omega_k} d\mu (\lambda) - \int_{\Omega_k - \Omega_k^*} d\mu (\lambda) \right).$$

(44)

On the other hand, according to the power constraint, we have

$$0 \leq \sum_{k=1}^{L_T} \int_{\Omega_k} d\mu (\lambda) - \sum_{k=1}^{L_T} \int_{\Omega_k} d\mu (\lambda)$$

$$= \sum_{k=1}^{L_T} \left( \int_{\Omega_k^* - \Omega_k} d\mu (\lambda) - \int_{\Omega_k - \Omega_k^*} d\mu (\lambda) \right).$$

(45)

Substituting (45) into (44) yields $I^* - I \geq 0$. As this holds for arbitrary $P_{on} \in \mathbb{R}^+$ the proof is complete.
C. Proof of Theorem 3

Theorem 3 rests on the following observation.

Lemma 4: For a differentiable function \( h(x) \) defined on \((a,b)\), denote its first derivative \( h'(x) \). If \( h(x) = 0 \) implies \( h'(x) < 0 \), then \( h(x) \) has at most one zero in its domain. If it exists, denote this unique zero by \( x_0 \). Then \( h(x) > 0 \) for all \( x \in (a,x_0) \) and \( h(x) < 0 \) for all \( x \in (x_0,b) \).

Proof: We first show that \( h(x) \) has at most one zero. For any zero \( x_0 \) of \( h(x) \), \( h'(x_0) < 0 \) by assumption and so there is an \( \epsilon > 0 \) such that \( h(x_0 + \epsilon) < 0 \), \( h(x_0 - \epsilon) > 0 \) and \( h(x) \neq 0 \) for all \( x \in (x_0 - \epsilon, x_0 + \epsilon) \) but \( x_0 \). Now suppose that \( x_1 \in (a,b) \) is another zero of \( h(x) \); we can assume that \( x_1 > x_0 \). Then \( x_0 < x_1 + \epsilon/2 < x_1 \). As \( h(x) \) is continuous, \( h(x) \) crosses the \( x \) axis at \( x_1 \) from negative to positive as \( x \) increases. But this would imply \( h'(x_1) > 0 \), which is a contradiction. Now with \( x_0 \) the unique zero (if it exists), that \( h(x) > 0 \) for all \( x \in (a,x_0) \) and \( h(x) < 0 \) for all \( x \in (x_0,b) \) follows by continuity.

Next recall that \( \overline{T}_\infty(a) \) can be expressed as the definite integral in (17). Let

\[
J = 1 - \log(1 + z(a)) - \int_0^a \frac{1}{1 + z(t)} \cdot \frac{f_r(t)}{\frac{c_r}{\infty}} dt. \quad (46)
\]

The sign of \( \overline{T}_\infty(a) \) is uniquely determined by that of \( J \) as the term \( f_r(a) \) in (17) is positive for all \( a \in (0,\pi) \).

We will show that \( J(a) = 0 \) implies \( J'(a) < 0 \). Note that

\[
J'(a) = -\frac{\frac{c_r}{\infty} z(a) + \frac{c_{r_{\infty}}}{\infty} 2\sqrt{\gamma} \sin(a)}{1 + z(a)} + \frac{f_r(a)}{\frac{c_r}{\infty}} - \frac{f_r(a)}{\frac{c_r}{\infty}} \int_0^a \frac{f_r(t)}{(1 + z(t))^2} \frac{c_r}{\infty} dt
\]

\[
= -\frac{f_r(a)}{\frac{c_r}{\infty}} \left[ \frac{z(a) - 1}{z(a) + 1} + \int_a^\pi \frac{f_r(t)}{(1 + z(t))^2} \frac{c_r}{\infty} dt \right] - \frac{\frac{c_r}{\infty} 2\sqrt{\gamma} \sin(a)}{1 + z(a)}.
\]

(47)

Since when \( J = 0 \) we have

\[
1 - \log(1 + z(a)) = \int_a^\pi \frac{1}{1 + z(t)} \frac{f_r(t)}{\frac{c_r}{\infty}} dt
\]

the second term in (47) is bounded below as in

\[
\int_a^\pi \frac{f_r(t)}{(1 + z(t))^2} \frac{c_r}{\infty} dt \geq \left( \int_a^\pi \frac{f_r(t)}{1 + z(t)} \frac{c_r}{\infty} dt \right)^2 \geq (1 - \log(1 + z(a)))^2
\]

where we have used Jensen’s inequality. So, the entire expression in the square brackets of (47) is bounded below by

\[
\frac{z(a) - 1}{z(a) + 1} + \int_a^\pi \frac{f_r(t)}{(1 + z(t))^2} \frac{c_r}{\infty} dt
\]

\[
\geq \left( \frac{z(a) - 1}{z(a) + 1} + (1 - \log(1 + z(a)))^2 \right) > 0
\]

where the last inequality follows as \( z(a) > 0 \) for \( a \in (0,\pi) \) and the easily verified fact that \( \frac{z(a) - 1}{z(a) + 1} + (1 - \log(1 + x))^2 > 0 \) for \( x > 0 \). Hence, the first part of (47) is negative; it is also true that the last term in (47) is always negative for \( a \in (0,\pi) \).

Now suppose in accordance with Lemma 4, \( J \) has a unique zero \( a_0 \in (0,\pi) \), and \( J > 0 \) for \( 0 < a < a_0 \) and \( J < 0 \) for \( a_0 < a < \pi \). Since the signs of \( \overline{T}_\infty(a) \) and \( J(a) \) are the same, the same conclusion holds for \( \overline{T}_\infty(a) \). In this case, \( \overline{T}_\infty \) is maximized at \( a_0 \in (0,\pi) \). If instead \( \overline{T}_\infty(a) \) does not have a root in \((0,\pi)\), note that as \( a \to \pi \), we have \( z(a) \to +\infty \), \( \log(1 + z(a)) \to +\infty \), and \( J \to -\infty \). Then \( J < 0 \) for \( a \in (0,\pi) \). Hence, \( \overline{T}_\infty(a) < 0 \) for all \( a \in (0,\pi) \) and \( \overline{T}_\infty \) is maximized at \( a = 0 \). The theorem is proved.

D. Proof of Corollary 1

The proof follows the same line as that for Theorem 3 (see Appendix C). Let \( J \) be as in (46). Then, the \( a_0 \) that maximizes \( \overline{T}_\infty \) should be either the unique zero of \( J \) if it exists, or \( 0 \) if \( J \) has no zero in \((0,\pi)\). We first prove that \( J = 0 \) implies \( \frac{dJ}{d\rho} < 0 \) for a given \( a \in (0,\pi) \) and \( \rho > 0 \). Then we show that \( a_0 \) is a nondecreasing function of \( \rho \).

For a given \( a \in (0,\pi) \) and \( \rho > 0 \), \( J = 0 \) implies \( \frac{dJ}{d\rho} < 0 \). Recall \( z(t) \) from (18); this is a function of \( \rho \). Computing \( \frac{dJ}{d\rho} \) gives

\[
\frac{dJ}{d\rho} = -\frac{1}{\rho} \left[ z(a) \int_0^\pi \frac{z(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt \right] - \int_0^\pi \frac{z^2(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt
\]

\[
= -\frac{1}{\rho} \left[ z(a) \int \frac{z^2(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt \right] + \int_0^\pi \frac{z^2(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt
\]

\[
\frac{dJ}{d\rho} = \int_0^\pi \frac{z(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt - \frac{z(a)}{1 + z(a)} \frac{f_r(t)}{\frac{c_r}{\infty}} dt < 1
\]

since \( \frac{z(t)}{1 + z(t)} < 1 \). This implies \( z(a) < e - 1 \). Furthermore

\[
\int_0^\pi \frac{z^2(t)}{(1 + z(t))^2} \frac{f_r(t)}{\frac{c_r}{\infty}} dt \geq \frac{z^2(a)}{(1 + z(a))^2}
\]

on account of the fact that \( \frac{z(t)}{1 + z(t)} \geq \frac{z(a)}{1 + z(a)} \) for all \( t \in (a,\pi) \). Now use the elementary inequality

\[
\frac{z(t)}{1 + z(t)} \geq \log(1 + x) + \left( \frac{x}{1 + x} \right)^2 > 0 \text{ valid for } 0 < x < e - 1 \text{ and substitute (49) and (50) into (48) to find } \frac{dJ}{d\rho} < 0 \text{ for } a \in (0,\pi) \text{ whenever } J = 0.
\]

Next let \( a_0 \) maximize \( \overline{T}_\infty \) for an SNR \( \rho_0 > 0 \), and let \( a_1 \) maximize \( \overline{T}_\infty \) for an SNR \( \rho_1 > 0 \) with \( \rho_0 < \rho_1 \). We will show that \( a_1 \leq a_0 \).
First take the case that \( a_0 > 0 \). We have \( J|_{a_0,p_0} = 0 \) by Theorem 3. Since \( J = 0 \) implies \( \frac{dJ}{dp} < 0 \) for \( a = a_0 \), \( J|_{a_0,p_0} < 0 \) by Lemma 4 in Appendix C. But \( a_1 \) maximizes \( J_{\infty}\) at \( p_1 \). Then either \( a_1 = 0 \), or \( a_1 > 0 \) and \( J|_{a_1,p_1} = 0 \) again by Theorem 3. If \( a_1 = 0 \), \( a_1 < a_0 \). If \( a_1 > 0 \), then \( \frac{dJ}{dp} |_{0 < \alpha \leq a_1, \varphi_1 > 0} > 0 \) and \( \frac{dJ}{dp} |_{a_1 < \alpha \leq a_0, \varphi_1 > 0} < 0 \) according to the proof in Appendix C. Since we have shown \( J|_{a_0,p_0} < 0, a_1 < a_0 \). That is, \( a_1 < a_0 \) if \( a_0 > 0 \). Next take \( a_0 = 0 \) and suppose that \( a_1 > a_0 \), then \( a_1 \in (0, \pi) \), \( J|_{a_1,p_1} = 0 \) and \( J|_{a_1,p_1} > 0 \). Because \( J|_{\alpha,\varphi} \to \infty \), there is an \( a' \in (a_1, \pi) \) such that \( J|_{a',\varphi} = 0 \). According to Theorem 3, \( a' \) maximizes \( J_{\infty} \) for \( \rho_0 \), contradicting the assumption that \( a_0 = 0 \) maximizes \( J_{\infty} \) for \( \rho_0 \). Therefore, \( 0 \leq a_1 \leq a_0 = 0 \) and, thus, \( a_1 = a_0 = 0 \).

Last, note that \( J_{\infty} = \int_{s} J_{\infty}(t) dt \) is a decreasing function of \( a \), and so \( J_{\infty} \) is increasing with \( \rho \). This completes the proof.

E. Proof of Theorem 6

Within a subcode with rank \( \delta(H) \), the optimal index is clearly given by

\[
\hat{\delta}(H) = \text{arg max}_{\delta: Q_{\bar{H}}(H)} \log |L_{L_R} + P_{\text{CHQ}} H^T Q_{\bar{H}}^T|.
\]

Thus the only nontrivial part is to prove the optimality of

\[
\hat{\delta}(H) \triangleq \max \left\{ s : J_{\hat{\delta}}(H) - I_{\hat{\delta}}(H) \geq (s - t) \kappa, 0 \leq t < s \right\}
\]

where \( J_{\hat{\delta}}(H) \) is defined in (39). For this we require the following.

**Lemma 5:** For all \( t > \tilde{s} \), \( I_{t}(H) - J_{\hat{\delta}}(H) < (t - \tilde{s}) \kappa \).

**Proof:** Suppose to the contrary that there is a \( t > \tilde{s} \) such that \( I_{t}(H) - J_{\hat{\delta}}(H) \geq (t - \tilde{s}) \kappa \). Let \( t_0 > \tilde{s} \) be the minimum \( t \) such that \( I_{t}(H) - J_{\hat{\delta}}(H) \geq (t - \tilde{s}) \kappa \). Then, if \( 0 \leq t \leq \tilde{s} \),

\[
I_{t_0} - I_{\hat{\delta}} = I_{t_0} - J_{\hat{\delta}} + J_{\hat{\delta}} - I_{\hat{\delta}} \\
\geq (t_0 - \tilde{s}) \kappa + (\tilde{s} - t) \kappa \\
= (t_0 - t) \kappa.
\]

where the inequality follows from the definitions of \( \tilde{s} \) and \( t_0 \). On the other hand, if \( \tilde{s} < t < t_0 \), \( I_{\tilde{s}} - I_{\hat{\delta}} \geq (t - \tilde{s}) \kappa \) because \( t_0 \) is the minimum integer such that \( I_{t}(H) - J_{\hat{\delta}}(H) \geq (t - \tilde{s}) \kappa \). Then

\[
I_{t_0} - I_{\hat{\delta}} = I_{t_0} - I_{\tilde{s}} + I_{\tilde{s}} - I_{\hat{\delta}} \\
\geq (t_0 - \tilde{s}) \kappa - (t - \tilde{s}) \kappa \\
= (t_0 - t) \kappa.
\]

Hence, \( I_{t_0} - I_{\hat{\delta}} \geq (t_0 - t) \kappa \) for all \( 0 \leq t \leq t_0 \), contradicting the definition of \( \tilde{s} \). The lemma is proved.

To prove \( \tilde{s} \) is optimal, we compare \( \hat{\delta}(\cdot) \) with an arbitrary deterministic feedback function \( \delta'(\cdot) \) which satisfies the power constraint. Let \( \delta' = \text{rank}(Q_{\delta'(H)}(H)) \) be the number of on-beams according to the feedback function \( \delta'(\cdot) \). Denote the distribution of \( H \) by \( \mu(H) \). Then

\[
\int \delta'(H) \mu(H) = \frac{P}{P_{\text{CH}}},
\]

Define \( \Delta s \triangleq \tilde{s} - \delta' \), and

\[
\Omega_{\Delta s} = \{ H \in \mathbb{C}^{L_R \times L_T} : \tilde{s} - \delta' = \Delta s \}.
\]

Then

\[
\bigcup_{\Delta s = L_T} \Omega_{\Delta s} = \mathbb{C}^{L_T \times L_s}.
\]

Since both \( \hat{\delta}(\cdot) \) and \( \delta'(\cdot) \) satisfy the power constraint, we have

\[
\int_{\Omega_{\Delta s}} \Delta s \cdot d\mu(H) = 0. \tag{51}
\]

The difference in information rate is lower bounded by

\[
\int_{\Omega_{\Delta s}} \left( I_{\delta}(H) - I_{\Delta s}(H) \right) d\mu(H)
\]

(b) follows from Lemma 5 and the definition of \( \delta \), and (c) follows from (51). Therefore, \( \hat{\delta}(\cdot) \) is the optimal feedback function and the proof is complete.

**REFERENCES**

[1] I. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[2] Physical Layer Aspects of Ultra High Speed Downlink Packet Access 3GPP TR25.848, 0.4.0 ed, 2000.

[3] V. K. N. Lau, L. Youjian, and T. A. Chen, “Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback,” *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2038–2049, 2004.

[4] V. Lau, L. Youjian, and T. A. Chen, “On the design of MIMO block-fading channels with feedback-link capacity constraint,” *IEEE Trans. Commun.*, vol. 52, pp. 62–70, 2004.

[5] C. Seong Taek and A. J. Goldsmith, “Degrees of freedom in adaptive modulation: A unified view,” *IEEE Trans. Commun.*, vol. 49, pp. 1561–1571, 2001.

[6] P. S. Chow and J. M. Cioffi, “Bandwidth optimization for high speed data transmission over channels with severe intersymbol interference,” in *Proc. IEEE Global Telecommun. Conf. (BLOBECOM)*, Orlando, FL, 1992, vol. 1, pp. 59–63.

[7] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, “On beamforming with finite rate feedback in multiple-antenna systems,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.
8. P. Xia, S. Zhou, and G. Giannakis, “Multiantenna adaptive modulation with beamforming based on bandwidth-constrained feedback,” IEEE Trans. Commun., vol. 53, pp. 526–536, 2005.

9. S. Zhou, Z. Wang, and G. 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.

10. J. C. Roh and B. D. Rao, “Vector quantization techniques for multiple-antenna channel information feedback,” in Proc. Int. Conf. Signal Process. Commun. (SPCOM), Bangalore, India, Dec. 11-14, 2004, pp. 402–406.

11. 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, 2003.

12. J. C. Roh and B. Rao, “Performance bounds for limited feedback MIMO beamforming systems,” in Proc. Allerton Conf. Commun., Control, Comput., Monticello, IL, Sep. 1, 2004.

13. R. S. Blum and J. H. Winters, “On optimum MIMO with antenna selection,” IEEE Commun. Lett., vol. 6, pp. 322–324, 2002.

14. W. Dai, Y. Liu, D. H. Cho, and E. A. Jorswieck, “An efficient feedback method for MIMO systems with slow, time-varying channels,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Atlanta, GA, Mar. 21-25, 2004, vol. 2, pp. 760–764.

15. D. Love and R. Heath, “Limited feedback unitary precoding for spatial multiplexing systems with linear receivers,” in Proc. Allerton Conf. Commun., Control, Comput., Monticello, IL, Oct. 1-3, 2003, pp. 685–694.

16. B. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, “Systematic design of unitary space-time constellations,” IEEE Trans. Inf. Theory, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

17. J. C. Roh and B. D. Rao, “Orthonormal basis expansion on the Grassmann manifold,” IEEE Trans. Inf. Theory, vol. 48, no. 9, pp. 2450–2454, 2002.

18. J. C. Roh and B. D. Rao, “MIMO spatial multiplexing systems with limited feedback,” in Proc. IEEE Int. Conf. Commun. (ICC), Seoul, Korea, May 17-21, 2005, pp. 777–782.

19. W. Dai, Y. Liu, D. H. Cho, and G. Giannakis, “Limited feedback unitary precoding for orthogonal space-time block codes,” IEEE Trans. Signal Process., vol. 53, pp. 64–73, 2005.

20. D. Love and R. Heath, Jr. and L. W. Hanlen, “Quantization on the Grassmann manifold: Applications to precoded MIMO wireless systems,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Philadelphia, PA, Mar. 18-23, 2005, pp. 1025–1028.

21. A. Barg and D. Y. Y. Ong, “On quantization of Grassmann manifolds,” IEEE Trans. Inf. Theory, vol. 48, no. 9, pp. 4134–4151, Dec. 2002.

22. W. Santipach and M. L. Honig, “Benefits of limited feedback for wireless channels,” in Proc. Allerton Conf. Commun., Control, Comput., Monticello, IL, Oct. 1-3, 2003.

23. W. Santipach and M. L. Honig, “Asymptotic performance of MIMO wireless channels with limited feedback,” in Proc. Military Commun. Conf., Boston, MA, Oct. 13–16, 2003, vol. 1, pp. 141–146.

24. W. Dai, Y. Liu, and B. Rider, “Quantization bounds on Grassmann manifolds and applications to MIMO systems,” IEEE Trans. Inf. Theory, vol. 54, pp. 1108–1123, Mar. 2008.

25. J. S. Serfiser, “Strong convergence of the empirical distribution of eigenvalues of large random matrices,” J. Multivar. Anal., vol. 55, no. 2, pp. 331–339, 1995.

26. A. Guionnet and O. Zeitouni, “Concentration of the spectral measure for large matrices,” Electr. Commun. Probab., vol. 5, pp. 119–136, 2000.

27. W. Dai, Y. Liu, and B. Rider, “Effect of finite rate feedback on CDMA signature optimization and MIMO beamforming vector selection,” IEEE Trans. Inf. Theory, vol. 55, no. 8, pp. 3651–3668, Aug. 2009.

28. J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc., in Grassmannian spaces,” Exper. Math., vol. 5, pp. 139–159, 1996.

29. Z. Llizhong and D. N. C Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” IEEE Trans. Inf. Theory, vol. 48, no. 2, pp. 359–383, 2002.

30. J. M. Kahn, Aspects of Multivariate Statistical Theory. New York: Wiley, 1982.

31. A. T. James, “Normal multivariate analysis and the orthogonal group,” Ann. Math. Statist, vol. 25, no. 1, pp. 40–75, 1954.