Interleaving Channel Estimation and Limited Feedback for Point-to-Point Systems with a Large Number of Transmit Antennas

Erdem Koyuncu, Xun Zou, and Hamid Jafarkhani

Abstract

We introduce and investigate the opportunities of multi-antenna communication schemes whose training and feedback stages are interleaved and mutually interacting. Specifically, unlike the traditional schemes where the transmitter first trains all of its antennas at once and then receives a single feedback message, we consider a scenario where the transmitter instead trains its antennas one by one and receives feedback information immediately after training each one of its antennas. The feedback message may ask the transmitter to train another antenna; or, it may terminate the feedback/training phase and provide the quantized codeword (e.g., a beamforming vector) to be utilized for data transmission. As a specific application, we consider a multiple-input single-output system with t transmit antennas, a short-term power constraint P, and target data rate ρ. We show that for any t, the same outage probability as a system with perfect transmitter and receiver channel state information can be achieved with a feedback rate of R_1 bits per channel state and via training R_2 transmit antennas on average, where R_1 and R_2 *are independent of* t, and depend only on ρ and P. In addition, we design variable-rate quantizers for channel coefficients to further minimize the feedback rate of our scheme.

Index terms: Interleaving, limited feedback, training, beamforming, partial CSIT and CSIR.

I. Introduction

The performance of a wireless communication system can be greatly improved by making the channel state information (CSI) available at the transmitter and the receiver. In a massive multiple-input single-output (MISO) system, having CSI at the transmitter (CSIT) is especially

This work was supported in part by the NSF Award ECCS-1611575.

This work was presented in part at the IEEE International Symposium on Information Theory (ISIT) in July 2015 [1].

E. Koyuncu is with Department of Electrical and Computer Engineering, University of Illinois at Chicago (e-mail: ekoyuncu@uic.edu). X. Zou and H. Jafarkhani are with Center for Pervasive Communications & Computing, University of California, Irvine (e-mails: xzou4@uci.edu; hamidj@uci.edu).

desirable as one can then fully exploit the performance gains promised by the large number of transmit antennas via CSI-adaptive transmission strategies such as beamforming. A typical way to acquire CSIT is channel estimation followed by (digital) feedback.

Channel training/estimation and feedback are traditionally viewed as two non-interleaving processes, as shown in Fig. 1. According to this traditional viewpoint, for each channel state, the transmitter first trains all of its antennas at once, so that the receiver acquires the entire CSI (or, in general, an erroneous version thereof.). This initial training phase is followed by the receiver feeding back a possibly-quantized version of the CSI. The receiver's feedback is then utilized at the transmitter side for data transmission (e.g., as a quantized beamforming vector.). Designing such limited feedback systems is a fundamental problem of communication theory and has been the subject of many publications [2]. In particular, limited feedback beamforming [3] has been studied through several different approaches that utilize Grassmannian line packings [4], vector quantization [5], combinations with orthogonal [6] or quasi-orthogonal [7] space-time codes, variable-length coding [8], or other systematic constructions [9]. Conditions to achieve full diversity in a finite feedback scheme has been discussed in [10], [11]. Various distributed limited feedback schemes [12]–[16] provide generalizations to multi-user networks.

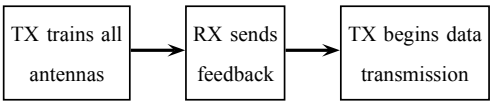

Fig. 1: Conventional training and limited feedback. TX and RX stand for the transmitter and the receiver, respectively.

The conventional scheme in Fig. 1 appears to be infeasible in the case of a massive MISO system. Even the channel training/estimation phase, by itself, would be very challenging to realize due to the large number of transmit antennas that need to be trained. Moreover, even if one assumes that the training stage somehow comes with no cost, feeding back the associated large number of channel values to the transmitter appears to be infeasible. Conventional limited feedback schemes also do not provide much hope in this context: The feedback rates required for even the simplest of the limited feedback schemes such as antenna selection grow without bound as the number of transmit antennas grows to infinity. In [17], it is analyzed in detail how many antennas per user terminal are needed to achieve some percentage of the ultimate performance limit with infinitely many antennas.

There has been some work on channel estimation and CSI feedback in massive MIMO systems; a survey can be found in [18]. In particular, [19] proposes a noncoherent trellis-coded quantization scheme, whose encoding complexity scales linearly with the number of antennas. In [20]–[22], compressive sensing techniques are utilized to reduce the feedback overhead of the CSI estimation. In addition, several studies [23]–[26] have demonstrated that channel or antenna correlation can be exploited to reduce the overhead of the downlink training phase. A multi-beam selection scheme for massive MIMO is presented in [27]. The problem of designing training sequences with low overheads have been studied in [28]. There are also several other approaches proposed for resolving the challenges of training and limited feedback in the more general context of multi-user MIMO; see e.g., [29]–[32].

Our proposed solution is to interleave the training and feedback stages as shown in Fig. 2. Unlike the conventional scheme in Fig. 1, the transmitter trains its antennas one by one and receives feedback information after training each one of its antennas. A feedback message may ask the transmitter to train another antenna (and also provide side information about the channel state), or it may result in the termination of the training phase, in which case it also provides the quantized codeword to be utilized by the transmitter for data transmission.

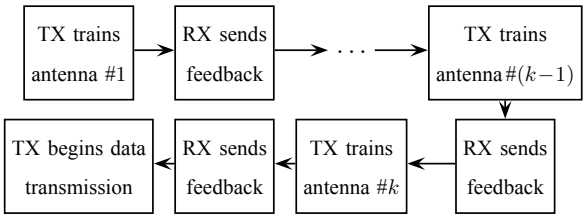

Fig. 2: Interleaved training and limited feedback. The number of trained antennas k varies from one channel state to another, and is itself decided through the training and feedback phases.

An interleaved scheme offers the following unique opportunity: If the already-trained antennas provide sufficiently favorable conditions for data transmission, one can then terminate the training phase and thus avoid wasting more resources on training the rest of the antennas. One main message of this paper is that in certain scenarios, we can make use of this opportunity to design multi-antenna communication systems whose feedback and training overheads remain completely independent of the number of transmit antennas, and which, at the same time, can achieve the same outage performance as a system with perfect transmitter and receiver CSI. Specifically, we

consider here a single-user point-to-point MISO system with the outage probability performance measure. Note that, while the "mainstream" use case of a massive transmitter antenna array is to support multiple users, a single-user system suffering from severe path loss may also greatly benefit from beamforming over a large number of antennas. Extensions to multiple-input multiple-output (MIMO) systems, or to multi-user scenarios with different performance measures (such as ergodic capacity) will thus be left as future work. In fact, after the publication of a preliminary version of this work [1], another paper [33] has studied the benefits of interleaving in hybrid single-user and multiple-user massive MIMO systems. The work [33] also considers a general channel model that can incorporate channel correlations. On the other hand, [33] ignores the feedback overhead of the interleaved scheme: It is assumed that the trained channel gains can be perfectly fed back to the transmitter, which requires an infinite number of feedback bits in practice. In contrast, we design interleaved schemes to minimize the training overhead as well as the feedback rate. In more detail, the main contributions of this paper are as follows:

- We propose a novel communication scheme which interleaves training and feedback stages. In this scheme, the transmitter trains its antenna one by one while the receiver transmits the feedback information immediately after training each antenna. The feedback message may ask the transmitter to train another antenna or provide the quantized codeword to be utilized for data transmission. The latter event occurs if the already trained antennas can provide enough channel gain to avoid outage.
- We apply the interleaving scheme to a MISO system with t transmit antennas, a short-term power constraint P, and target data rate ρ. We show that our scheme is able to achieve the same outage probability as a system with perfect transmitter and receiver CSI while keeping the average feedback rate and the average number of training antennas independent of t and dependent only on P and ρ.
- We design a variable-rate quantizer to minimize the feedback rate in the MISO system while keeping the same outage probability as a full-CSI system. It is achieved by allocating a higher rate to a larger coefficient in a given channel state. We also discuss the latency costs associated with interleaving, and study antenna grouping schemes as a solution.

Part of this work has been presented in a conference [1]. Compared to [1], the current paper provides the proofs of technical results. It also describes a procedure to design optimal variable-

TABLE I: Table of Symbols for Different Schemes.

Symbol	Definition
F	Full-CSI scheme
G	Open-loop scheme
A	Conventional antenna selection scheme
B	Interleaved antenna selection scheme
B′	Interleaved antenna selection scheme with antenna grouping
S	Conventional beamforming scheme
D	Interleaved beamforming scheme
C$_i$, $i = 1, \ldots, t$	Sub-blocks of interleaved beamforming scheme

rate quantizers of the feedback information. Here, we also provide numerical results that verify our analysis, and a discussion on the latency costs associated with interleaving and grouping.

The rest of the paper is organized as follows: In Section II, we describe the system model, and the full-CSI and open-loop systems. In Section III, we introduce the idea of interleaving and construct a simple interleaved scheme based on antenna selection. In Section IV, we show how to design an interleaved scheme that can achieve the full-CSI gains with low training length and low feedback rate. In Section V, we describe a variable-rate quantizer to further reduce the feedback rate. In Section VI, we discuss latency costs and study interleaved schemes with antenna grouping. Finally, we present the simulation results in Section VII and conclusions in Section VIII. Some of the technical proofs and extended discussions are provided in the appendices.

Notation: $\mathbb{C}^{m\times n}$ is the set of all $m \times n$ complex matrices with $\mathbb{C}^m \triangleq \mathbb{C}^{m\times 1}$ and $\mathbb{C} \triangleq \mathbb{C}^1$. $\mathbb{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{0}_{m\times n}$ is the $m \times n$ all-zero matrix. $\mathrm{CN}(\mathbf{K})$ is a circularly-symmetric complex Gaussian random vector with covariance matrix $\mathbf{K}$. P and E represent the probability and the expected value, respectively. o, O, and Θ are the standard Bachmann-Landau symbols. $\mathbf{A}^T$ and $\mathbf{A}^\dagger$ are the transpose and the conjugate transpose of matrix $\mathbf{A}$. $\circ$ stands for the entrywise product. For $x \in \mathbb{R}$, $x^+ \triangleq x$ if $x \geq 0$, and $x^+ \triangleq 0$, otherwise. For reader's convenience, we show the symbols that we will use for various schemes in the paper in Table I.

II. Preliminaries

We consider a MISO system with t transmit antennas. Denote the channel from transmit antenna i to the receiver antenna by h_i, and let $\mathbf{h} = [h_1 \cdots h_t]^T \in \mathbb{C}^t$ represent the entire channel state. We assume that $\mathbf{h} \simeq \mathrm{CN}(\mathbf{I}_t)$. The transmitted symbol $\mathbf{s} \in \mathbb{C}^t$ and the received

symbol $y \in \mathbb{C}$ have the input-output relationship $y = \mathbf{s}^T\sqrt{P}\mathbf{h} + \eta$, where P is the short-term power constraint of the transmitter, i.e., the total transmit power constraint over t antennas, and the noise term $\eta \sim \mathrm{CN}(1)$ is independent of $\mathbf{h}$.

For a fixed $\mathbf{h}$, suppose that input symbol $\mathbf{s}$ is distributed as $\mathrm{CN}(\mathbf{K}^T)$, where $\mathbf{K}$ is a covariance matrix with $\mathrm{tr}(\mathbf{K}) \leq 1$. With perfect CSIR, the channel capacity under this strategy is $\log_2(1 + \mathbf{h}^\dagger\mathbf{K}\mathbf{h}P)$ bits/sec/Hz. In this work, we consider a delay-constrained system where it is necessary and sufficient to sustain a certain fixed rate of data transmission at all times. Examples include video streaming for teleconferencing. In these so-called block-fading scenarios, averaging out a data codeword over infinitely many channel states is not feasible. The appropriate performance metric is the outage probability, which is the probability that the system will not be able to support a given target data rate [34], [35]. In our system, for a given target data transmission rate $\rho = \log_2(1 + \alpha P)$, where $\alpha > 0$ can be chosen arbitrarily, an outage event occurs if $\log_2(1 + \mathbf{h}^\dagger\mathbf{K}\mathbf{h}P) < \rho$, or equivalently if $\mathbf{h}^\dagger\mathbf{K}\mathbf{h} < \alpha$. We refer to the special case where $\mathbf{K} = \mathbf{x}\mathbf{x}^\dagger$ for some $\mathbf{x} \in \mathbb{C}^t$ with $\|\mathbf{x}\| \leq 1$ as "beamforming," in which case the outage event is $|\langle \mathbf{x}, \mathbf{h}\rangle|^2 < \alpha$. We assume that both the transmitter and the receiver agree upon a common transmission rate and power before any training or feedback communication takes place. This ensures that both terminals have perfect knowledge of α. We also assume that there are no CSI estimation errors: Once a transmitter trains a particular antenna, the receiver can acquire the corresponding CSI error-free. The results of this paper will thus serve as upper bounds on the performance of systems that take into account possible errors in CSI estimation.

For a random $\mathbf{h}$, the transmitter can use different covariance matrices for different $\mathbf{h}$. Let $\mathtt{M} : \mathbb{C}^t \to \mathbb{C}^{t\times t}$ be an arbitrary mapping, so that given $\mathbf{h}$, the input symbol is distributed as $\mathrm{CN}([\mathtt{M}(\mathbf{h})]^T)$. The outage probability with $\mathtt{M}$ is $\mathtt{out}(\mathtt{M}) \triangleq \mathrm{P}(\mathbf{h}^\dagger\mathtt{M}\mathbf{h} < \alpha)$. For a beamforming-only system with mapping $\mathtt{N} : \mathbb{C}^t \to \mathbb{C}^t$, we define $\mathtt{out}(\mathtt{N}) \triangleq \mathrm{P}(|\langle \mathtt{N}(\mathbf{h}), \mathbf{h}\rangle|^2 < \alpha)$.

With perfect CSIT and CSIR (a "full-CSI" system), the optimal mapping is beamforming along $\mathbf{h}$ [36]. In other words, the mapping $\mathtt{F}(\mathbf{h}) \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|}$ provides the minimum-possible outage probability $\mathtt{out}(\mathtt{F}) = \mathrm{P}(\|\mathbf{h}\|^2 \leq \alpha)$. With perfect CSIR but no CSIT (an "open-loop" system), it is shown in [38] that the optimal mapping is $\mathtt{G}(\mathbf{h}) \triangleq \frac{1}{\kappa}\begin{pmatrix} \mathbf{I}_\kappa & \mathbf{0}_{\kappa\times(t-\kappa)} \\ \mathbf{0}_{(t-\kappa)\times\kappa} & \mathbf{0}_{(t-\kappa)\times(t-\kappa)} \end{pmatrix}$, where $\kappa \triangleq \arg\min_k \mathrm{P}(\sum_{i=1}^k |h_i|^2 < k\alpha)$. Hence, only κ out of the t antennas are used in general, and we have $\mathtt{out}(\mathtt{G}) = \mathrm{P}(\|\mathbf{h}_\kappa\|^2 < \kappa\alpha)$. Note that κ does not depend on the channel state $\mathbf{h}$. Therefore, the open-loop mapping is also independent of the channel state.

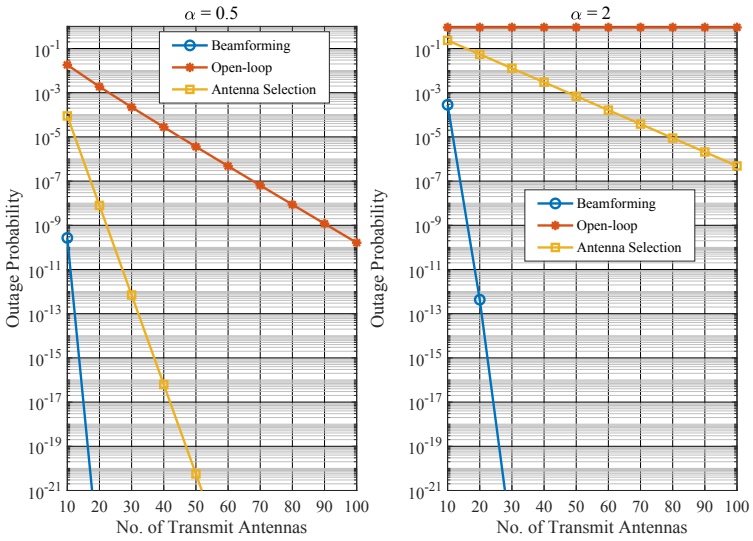

Fig. 3: Outage probability as a function of the number of transmit antennas t for beamforming, open-loop, and antenna selection schemes at $\alpha = 0.5$ (left) and $\alpha = 2$ (right). Note that a larger path loss exponent (due to higher frequency of transmission) or a greater transmitter-to-receiver separation translates to a higher α in practice. Therefore, using as many as a hundred antennas may be necessary to achieve an acceptable outage probability even in a single-user system, as evidenced by the case $\alpha = 2$ and antenna selection.

The outage performance of communication systems in terms of their α-asymptotic behaviors for a fixed t has been studied in the literature. For example, a full-CSI and an open-loop system, with t antennas, both provide a "diversity gain" of t [36]. In other words, given a fixed t, as $\alpha \to 0$, we have $\mathtt{out}(\mathtt{F}) \in \Theta(\alpha^t)$ and $\mathtt{out}(\mathtt{G}) \in \Theta(\alpha^t)$ so that the outage probabilities of a full-CSI and an open-loop system have the same $\alpha \to 0$ behavior. In contrast, in this work, we are primarily interested in the t-asymptotic behavior of outage probabilities for a fixed α, i.e., the behavior of the system for a massive number of antennas. The following proposition, whose proof can be found in Appendix A, provides a rough characterization in this context.

Proposition 1 *As $t \to \infty$, for a full-CSI system, we have $\mathtt{out}(\mathtt{F}) \in \Theta(\frac{\alpha^t}{t!})$, $\forall \alpha > 0$, whereas for an open-loop system, we have $\mathtt{out}(\mathtt{G}) \in \Theta\big(\frac{(t\alpha)^t e^{-\alpha t}}{t!}\big)$ if $0 < \alpha < 1$, and $\mathtt{out}(\mathtt{G}) \in \Theta(1)$ if $\alpha \geq 1$.*

As is shown in Fig. 3, the outage probability of an open-loop system decays much slower than that of a full-CSI system. Proposition 1 brings both good and bad news. The good news is that for a full-CSI system, one can transmit with an arbitrarily large data rate (by choosing a sufficiently large α) with a fixed power consumption P and zero outage as $t \to \infty$. The bad

news is that it is not always possible to do the same in an open-loop system: When $\alpha \geq 1$, the outage probability does not decay to 0 with increasing t, and in fact, it saturates to a certain non-zero value. Also, for $0 < \alpha < 1$, even though we have $\mathtt{out}(\mathtt{G}) \to 0$ as $t \to \infty$, there is still room for improvement: As t increases, the outage probability of a full-CSI system decays much faster than that of an open-loop system.

In order to obtain a vanishing outage probability as $t \to \infty$ for every α, one should thus utilize CSIT. The full-CSI system is impractical as it requires an "infinite" rate of feedback from the receiver to the transmitter. A more practical approach is to settle for quantized CSIT via finite-rate receiver feedback [3]. Another issue that is common to both a full-CSI and an open-loop system is the requirement of perfect CSIR, which may, by itself, not be feasible when t is large. In the following, we thus consider the design of partial CSIT, partial CSIR schemes that interleave the training and feedback processes as shown in Fig. 2.

III. Interleaved Training and Limited Feedback

We begin with a simple example of an interleaved scheme that is based on antenna selection. We first describe its conventional non-interleaved counterpart.

A. The Conventional Antenna Selection Scheme

A well-known partial-CSIT scheme is what we shall refer to as the "conventional" antenna selection scheme: Given $\mathbf{h}$, the transmitter first trains all of its antennas so that the receiver acquires the entire CSI. The receiver determines the antenna index $\tau \triangleq \arg\max_i |h_i|$ with the highest channel gain and sends $\lceil \log_2 t \rceil$ feedback bits to the transmitter that can uniquely represent τ. The transmitter recovers τ from the feedback bits and transmits over antenna τ.

This scheme can be characterized by the mapping $\mathtt{A}(\mathbf{h}) \triangleq \mathbf{e}_\tau$, where $\mathbf{e}_i = [\mathbf{0}_{1\times(i-1)}\ 1\ \mathbf{0}_{1\times(t-i)}]^T$, $i = 1, \ldots, t$ are the standard basis vectors for $\mathbb{C}^t$. We have $\mathtt{out}(\mathtt{A}) = (1 - e^{-\alpha})^t$, which implies $\forall \alpha > 0$, $\lim_{t\to\infty} \mathtt{out}(\mathtt{A}) = 0$. Hence, for every $\alpha > 0$, we can obtain a vanishing outage probability as $t \to \infty$, as desired, which is also shown in Fig. 3. Moreover, for any α and t, we have $\mathtt{out}(\mathtt{A}) \leq \mathtt{out}(\mathtt{G})$, and in fact, it can be shown (e.g. by applying Stirling's approximation to the asymptotic formulae in Proposition 1) that $\mathtt{out}(\mathtt{A}) \in o(\mathtt{out}(\mathtt{G}))$, $\forall \alpha \in (0, 1)$. Hence, relative to an open-loop system, antenna selection improves the t-asymptotic behavior of the outage probability for all $\alpha > 0$. This is shown for two values of α in Fig. 3. On the other hand, to

implement this scheme, one needs to train t scalar channels (one for each h_i) and feed back $\lceil \log_2 t \rceil$ bits for every channel. Clearly, this is not feasible in the $t \to \infty$ regime.

B. A New Antenna Selection Scheme

The conventional antenna selection scheme is excessively precise in the sense that it always tries to select the antenna with the highest gain. On the other hand, without any loss of optimality in terms of the outage probability, we can in fact select *any* one of the antennas that avoids outage (not necessarily *the* antenna that provides the highest channel gain) whenever there is one. We use this observation to design an alternate antenna selection scheme that is based on the idea of interleaving training and limited feedback.

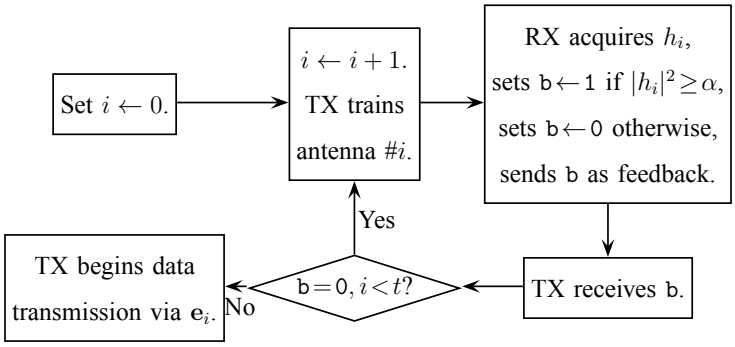

Fig. 4: The new antenna selection scheme. Note that the variable i, i.e., the antenna number, can be thought to be "naturally available" to both the transmitter and the receiver: At both terminals, it can be initialized and updated throughout the multiple training and feedback stages without any extra overhead. Also note that the receiver is always aware of what the next action (training a new antenna or beginning the data transmission) of the transmitter is going to be so that there is no inconsistency. This is because it is the receiver itself that provides the feedback message, which uniquely determines the transmitter action.

Our new antenna selection scheme operates as shown in Fig. 4: The transmitter first trains the channel h_1 corresponding to the first antenna and waits for receiver feedback. The receiver, having acquired the knowledge of h_1, sends the one-bit feedback message "1" if $|h_1|^2 \geq \alpha$, i.e. if selecting the first antenna avoids outage. Otherwise, it feeds back a "0," which indicates that selecting the first antenna will result in an outage. Now, if the transmitter receives a "1," the training and feedback process can end; the transmitter starts data transmission over the first antenna only (without the need of training the remaining antennas) and outage is avoided. Otherwise, if the transmitter receives a "0," it proceeds to training the channel state h_2

corresponding to its second antenna. The process continues in the same manner until an antenna (selection vector) that avoids outage is found. If all the antennas result in an outage, then the transmitter can simply transmit over an arbitrary antenna.

Clearly, the new scheme achieves the same outage probability $(1-e^{-\alpha})^t$ as the conventional scheme discussed in Section III-A. Now, given $1 \leq i \leq t-1$, the transmitter trains only the first i antennas with probability $e^{-\alpha}(1-e^{-\alpha})^{i-1}$, and it trains all the t antennas with probability $(1-e^{-\alpha})^{t-1}$. The training length, which we define as the average number of antennas that are trained per channel state, is thus

$$\sum_{i=1}^{t} ie^{-\alpha}(1-e^{-\alpha})^{i-1} + t(1-e^{-\alpha})^t = e^{\alpha}(1-(1-e^{-\alpha})^t).$$

A similar calculation reveals that the feedback rate of the scheme, which we define as the average number of bits that are fed back per channel state, is actually (numerically) equal to its training length. Hence, the training and the feedback rates of the new scheme are both given by the formula $e^{\alpha}(1-(1-e^{-\alpha})^t)$. Note that for any t, the two rates are both upper bounded by e^{α}, which is independent of t.

The significance of the new scheme is that it provides a vanishing outage probability as $t \to \infty$ with t-independent training length and feedback rate. One can thus obtain the benefits of having infinitely many antennas with finite training and feedback overheads. For example, setting $\alpha = 1$, we can observe that if the transmitter has infinitely many antennas, then for any given power constraint P, we can transmit with rate $\log(1+P)$ bits/sec/Hz outage-free via training only $e < 3$ antennas and feeding back 3 bits on average. Comparison with an open-loop system (a system with perfect CSIR but no CSIT) leads to the following conclusion: It is much better to have a little bit of CSIT and a little bit of CSIR rather than to have perfect CSIR but no CSIT.

We note that our interleaved antenna selection scheme can also be applied to the orthogonal frequency division multiplexing (OFDM) systems. The main challenge is that the best selection of antennas is likely to change with frequency. As is shown in [37], the antenna selection problem can be formulated as finding the antenna with the best channel averaged over all sub-carriers. As a result, we may use the average channel gain over all sub-carriers to determine whether a specific antenna is outage-avoiding or not.

Several variations on our interleaved antenna selection scheme can be considered. For example, in order to avoid the possible implementation complexities and delays of training the antennas

one by one, the transmitter may train all t antennas at once as in conventional antenna selection. On the other hand, the receiver may now use variable-length feedback instead of the $\lceil \log_2 t \rceil$ bits of fixed length feedback in conventional antenna selection. In detail, suppose that selecting any of the first υ antennas results in an outage, but selecting Antenna $\upsilon + 1$ avoids outage, where $\upsilon \in \{0, \ldots, t\}$. We let $\upsilon = t$ if selecting any of the t antennas results in outage. The receiver then feeds back the binary codeword $1 \cdots 10$, where there are υ ones. The transmitter can recover the outage-avoiding antenna from the feedback information if such an antenna exists. This scheme, which utilizes fixed-length training and variable-length feedback, lies in between the two extremes of conventional antenna selection (that uses fixed-length training and feedback), and interleaved antenna selection (that uses variable-length training and feedback). It is a special case of the variable-length beamforming schemes in [8] for full-CSIR systems. It achieves the same outage probability as conventional antenna selection with training length t, and feedback rate $e^{\alpha}(1 - (1 - e^{-\alpha})^t)$. Note that the feedback rate of the scheme equals that of interleaved antenna selection and thus remains bounded as $t \to \infty$. As discussed in [8], the feedback rate may possibly be reduced further with a better codeword assignment; e.g., by using Huffman's algorithm. Nevertheless, the training length of the scheme grows without bound as $t \to \infty$. Later in Section VI, we shall consider other variations that rely on training a subset of antennas at a time instead of training all antennas at once or training them one by one.

C. General Description of an Interleaved Scheme

So far, we have discussed many seemingly-different scenarios including non-interleaved or interleaved schemes, the full-CSI and the open-loop systems, and so on. All of these scenarios can in fact be viewed as manifestations of a single unifying framework of a *generalized beamforming scheme*, which describes the rules of how the tasks of training and feedback are to be performed. The advantage of this viewpoint is that it will allow us to more meaningfully compare different scenarios with respect to their outage probabilities, training lengths, and feedback rates. We call this generalized beamforming scheme, as defined below, Scheme $\mathtt{S}$.

One task of Scheme $\mathtt{S}$ is to specify the quantized covariance matrix $\mathtt{S}(\mathbf{h})$ to be utilized given channel state $\mathbf{h}$. By the definitions in Section II, the outage probability with $\mathtt{S}$ is thus given by $\mathtt{out}(\mathtt{S})$. Scheme $\mathtt{S}$ also describes which antennas are to be trained in which order, the corresponding feedback messages of the receiver, and how these messages are decoded at the

transmitter. Obviously, different choices result in different schemes and different performances. An example of these "inner workings" of Scheme $\mathtt{S}$ can be found in Section III-B for the special case of our new antenna selection scheme. As such, while we use Scheme $\mathtt{S}$ to represent the general structure of our beamforming scheme, when the details of training, feedback, transmission and decoding are defined, i.e., a specific scheme is defined in details as done in Section III-B, we will use a specific name for the specific scheme. The two important figures of merit of Scheme $\mathtt{S}$ is its training length $\mathtt{tl}(\mathtt{S})$ and its feedback rate $\mathtt{fr}(\mathtt{S})$, which can be defined in the same manner as we have done in Section III-B.

We can now view a full-CSI system, called Scheme $\mathtt{F}$, as an example of Scheme $\mathtt{S}$. Operationally, a full-CSI system trains all its antennas and performs the optimal beamforming along the direction $\frac{\mathbf{h}}{\|\mathbf{h}\|}$. As a result, we will have $\mathtt{out}(\mathtt{F}) = \mathrm{P}(\|\mathbf{h}\|^2 < \alpha)$ and $\mathtt{tl}(\mathtt{F}) = t$. Since representing an arbitrary beamforming vector requires an infinite rate of feedback, we have $\mathtt{fr}(\mathtt{F}) = \infty$. Similarly, the open-loop scheme $\mathtt{G}$ trains the first κ antennas. Since there is no feedback, $\mathtt{fr}(\mathtt{G}) = 0$ and the transmitter sends independent Gaussian symbols with equal energy over the first κ antennas. Therefore, we have $\mathtt{out}(\mathtt{G}) = \mathrm{P}(\|\mathbf{h}_\kappa\|^2 < \kappa\alpha)$ and $\mathtt{tl}(\mathtt{G}) = \kappa$. Also, as shown in Section III-A, the conventional antenna selection system, called Scheme $\mathtt{A}$, will have $\mathtt{out}(\mathtt{A}) = (1 - e^{-\alpha})^t$, $\mathtt{tl}(\mathtt{A}) = t$, and $\mathtt{fr}(\mathtt{A}) = \lceil \log_2 t \rceil$.

Clearly, Scheme $\mathtt{S}$ provides a framework to extend the previous definitions in a consistent manner and offers a set of quantities to compare the performance of different schemes. For example, we can summarize the performance metrics of our new antenna selection scheme in Section III-B, called Scheme $\mathtt{B}$, in the following theorem:

Theorem 1 *Scheme* $\mathtt{B}$*, defined in Section III-B, provides* $\mathtt{out}(\mathtt{B}) = \mathtt{out}(\mathtt{A}) = (1 - e^{-\alpha})^t$ *and* $\mathtt{tl}(\mathtt{B}) = \mathtt{fr}(\mathtt{B}) = e^{\alpha}(1 - (1 - e^{-\alpha})^t) < e^{\alpha}$.

These results lead to the following question: What is the best-possible outage probability for given constraints on training length and feedback rate? Unfortunately, this problem appears to be difficult in general, and we thus leave a detailed treatment as future work. In a related direction, Theorem 1 shows the existence of a "good" scheme that can achieve a vanishing outage probability as $t \to \infty$ with t-independent feedback and training lengths. One fundamental question that immediately comes to mind is then to determine whether one can achieve the ultimate limit $\mathtt{out}(\mathtt{F})$ with again t-independent training length and feedback rate. The answer is

yes, and the construction of such a scheme will be provided next. Meanwhile, we note that even though antenna selection provides a reasonable performance, we still have $\mathtt{out}(\mathtt{F}) \in o(\mathtt{out}(\mathtt{A}))$ as $t \to \infty$. In other words, the outage probability with a full-CSI system decays much faster than the one with antenna selection. While we have shown this fact analytically, Fig. 3 demonstrates it numerically as well. This also provides a "practical motivation" for construction of schemes that achieve the full-CSI gains.

IV. Achieving the Full-CSI Gains by Interleaving

Our construction here relies on our earlier work [8], which introduced the idea of variable-length feedback for a MISO system with perfect CSIR. We thus first recall some of the relevant technical tools and results.

A. Variable-Length Limited Feedback with Perfect CSIR

We begin by defining a simple deadzone scalar quantizer. For any given integer $\ell \geq 0$ and $x \in [-1, +1]$, let $q(x;\ell) \triangleq \mathrm{sign}(x)\frac{1}{2^{\ell+1}}\lfloor |x| 2^{\ell+1} \rfloor$. We can easily calculate $q(x;\ell)$ by taking the most significant $\ell+2$ bits $(b_0.b_1b_2\cdots b_{\ell+1})_2$ of the binary representation $(b_0.b_1b_2\cdots)_2$ of $|x|$, while preserving the sign of x. For example, we have $q(\pm(0.101)_2;1) = \pm(0.10)_2$.

We extend the definition of the deadzone quantizer q to an arbitrary beamforming vector $\mathbf{x} = [x_1 \cdots x_t]^T \in \mathbb{C}^t$ with $\|\mathbf{x}\| \leq 1$ by setting $q(\mathbf{x};\ell) \triangleq [\, q(\Re x_1;\ell) + jq(\Im x_1;\ell) \;\cdots\; q(\Re x_t;\ell) + jq(\Im x_t;\ell)\,]^T \in \mathbb{C}^t$. We refer to the parameter ℓ as the "resolution" of q. Note that by construction, $\|q(\mathbf{x};\ell)\| \leq 1$, and therefore, $q(\mathbf{x};\ell)$ is itself a feasible beamforming vector. Moreover, for a fixed ℓ and t, each quantized vector $q(\mathbf{x};\ell)$ can be uniquely represented by $2t(\ell+3)$ bits (For each of the $2t$ complex dimensions of $\mathbf{x}$, we spend one bit for the sign, and $\ell+2$ bits for the most significant $\ell+2$ binary digits.).

Now, for an arbitrary channel state $\mathbf{h}$ with $\|\mathbf{h}\|^2 > \alpha$, let $L(\mathbf{h}) \triangleq \max\{\lceil \log_2(4t) \rceil, \lceil \log_2 \frac{4t\alpha}{\|\mathbf{h}\|^2-\alpha} \rceil\}$, and $\overrightarrow{\mathbf{h}} \triangleq \mathtt{F}(\mathbf{h}) = \frac{\mathbf{h}}{\|\mathbf{h}\|}$. We have the following proposition.

Proposition 2 ([8, Proposition 4]) *Let $\mathbf{h} \in \mathbb{C}^t$ with $\|\mathbf{h}\|^2 > \alpha$ for some $t \geq 1$. Then,*

$$|\langle q(\overrightarrow{\mathbf{h}}; L(\mathbf{h})), \mathbf{h}\rangle|^2 > \alpha. \tag{1}$$

This result has the following interpretation. Suppose $\|\mathbf{h}\|^2 > \alpha$, and thus outage is avoidable with the beamforming vector $\overrightarrow{\mathbf{h}}$. By construction, the sequence of quantized beamforming vectors

$q(\overrightarrow{\mathbf{h}};\ell)$, $\ell \geq 0$ (which are feasible since $\|q(\overrightarrow{\mathbf{h}};\ell)\| \leq \|\overrightarrow{\mathbf{h}}\| = 1$) provides an increasingly finer approximation of $\overrightarrow{\mathbf{h}}$ as the resolution ℓ grows to infinity. The proposition shows that for every given $\mathbf{h}$ with $\|\mathbf{h}\|^2 > \alpha$, there is in fact a "sufficient resolution" $L(\mathbf{h})$ (that depends only on $\|\mathbf{h}\|$) such that the quantized beamforming vector $q(\overrightarrow{\mathbf{h}};\ell)$ can avoid outage.

As discussed in [8], Proposition 2 leads to the following limited feedback scheme under the assumption of perfect CSIR: If $\|\mathbf{h}\|^2 > \alpha$, the receiver calculates the required resolution $L(\mathbf{h})$ to avoid outage, and sends $2t(L(\mathbf{h})+3)$ feedback bits that represent the corresponding outage-avoiding beamforming vector $q(\overrightarrow{\mathbf{h}};L(\mathbf{h}))$. The transmitter, which we assume can perfectly know the length of the feedback codeword that it has received, first recovers $L(\mathbf{h})$, and then the beamforming vector $q(\overrightarrow{\mathbf{h}};L(\mathbf{h}))$. Otherwise, if $\|\mathbf{h}\|^2 \leq \alpha$, outage is unavoidable except for channel states $\|\mathbf{h}\|^2 = \alpha$ with zero probability. In this case, the receiver sends the one-bit feedback message "0" so that the transmitter can transmit with an arbitrary but fixed beamforming vector, say $\mathbf{e}_1$. We refer to this scheme as Scheme $\mathtt{C}_t$, where the subscript indicates the number of transmit antennas. We have $\mathtt{C}_t(\mathbf{h}) = q(\overrightarrow{\mathbf{h}};L(\mathbf{h}))$. By construction, Scheme $\mathtt{C}_t$ achieves the full-CSI outage probability with the feedback rate

$$\mathtt{fr}(\mathtt{C}_t) = \mathrm{P}(\|\mathbf{h}\|^2 \leq \alpha) + \textstyle\sum_{\ell=\lceil \log_2(4t)\rceil}^{\infty} 2t(\ell+3)p_\ell, \tag{2}$$

where $p_\ell \triangleq \mathrm{P}(L(\mathbf{h})=\ell, \|\mathbf{h}\|^2 > \alpha)$. As $\ell \to \infty$, p_ℓ can be shown to decay fast enough so that the resulting feedback rate is finite; we refer the interested reader to [8] for the details and formal calculations. Intuitively, instead of trying to pick the best beamforming vector that maximizes the signal-to-noise ratio in some given codebook, one spends just enough bits to describe a beamforming vector that avoids outage. This allows us to achieve the full-CSI performance with a finite feedback rate under the assumption of perfect CSIR.

B. Achieving $\mathtt{out}(\mathtt{F})$ *by Interleaving*

We now return to our main goal of designing a scheme that can achieve the full-CSI outage probability with finite training length and feedback rate. Scheme $\mathtt{C}_t$ as described above is not immediately applicable for our purposes as (i) it requires perfect CSIR and thus induces a training length of t, and (ii) according to (2), its feedback rate grows at least as $\Theta(t)$ (We have $\mathtt{fr}(\mathtt{C}_t) \geq 6t \sum_{\ell=\lceil \log_2(4t)\rceil}^{\infty} p_\ell = 6t\mathrm{P}(\|\mathbf{h}\|^2 > \alpha) \in \Theta(t)$.).

We can however incorporate the sequence of Schemes $\mathtt{C}_i$, $i = 1,\ldots,t$ as sub-blocks of an interleaved training and limited feedback Scheme $\mathtt{D}$ as shown in Fig. 5. In the figure, we use the

notation $\mathbf{h}_i \triangleq [h_1 \cdots h_i]^T$, $i = 1, \ldots, t$ to represent the first i components of the channel state $\mathbf{h}$. Given $\mathbf{h}$ and a value of the variable $i \in \{1, \ldots, t\}$ in the figure, suppose that the transmitter has "just" trained its ith antenna, so that the receiver has acquired the knowledge of h_i. At this stage, the receiver knows the channel values $h_1, \ldots, h_i$ corresponding to the first i antennas of the transmitter, or equivalently, it knows $\mathbf{h}_i$. We consider the following two cases for the receiver's feedback and the corresponding transmitter action.

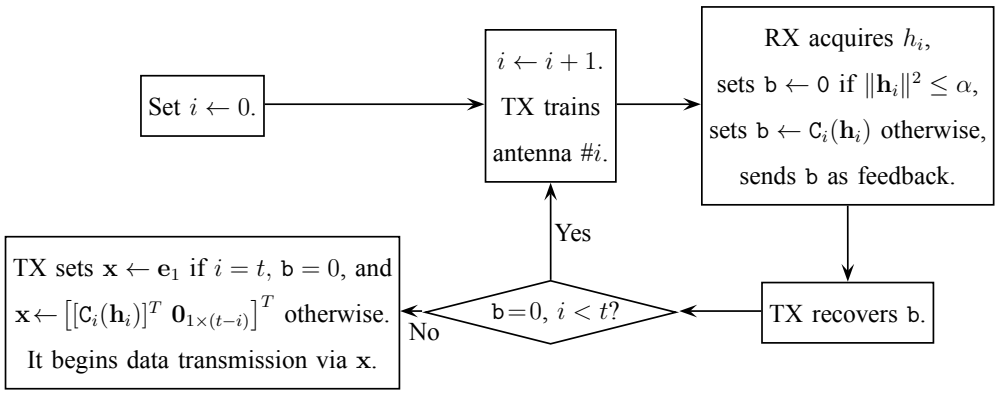

Fig. 5: Operation of scheme $\mathtt{D}$. Due to the equivalence between $\mathtt{C}_i(\mathbf{h}_i) = q(\overrightarrow{\mathbf{h}_i}; L(\mathbf{h}_i))$ and its binary description (see Section IV.A), we use the same notation "$\mathtt{C}_i(\mathbf{h}_i)$" for the codeword of $2i(L(\mathbf{h}_i)+3)$ bits that represent $\mathtt{C}_i(\mathbf{h}_i)$.

If $\|\mathbf{h}_i\|^2 \le \alpha$, as far as the channels that have been made available to the receiver are concerned, outage is unavoidable with probability 1. The receiver thus requests the transmitter to train the next antenna by sending the feedback bit "0," and the transmitter complies. The case $i = t$ is an exception: Outage is unavoidable with any beamforming vector with probability 1 (we have $\|\mathbf{h}_t\|^2 = \|\mathbf{h}\|^2 \le \alpha$), and thus the transmitter transmits via the (arbitrarily chosen) vector $\mathbf{e}_1$.

On the other hand, if $\|\mathbf{h}_i\|^2 > \alpha$, the receiver feeds back the i-dimensional vector $\mathtt{C}_i(\mathbf{h}_i) = q(\overrightarrow{\mathbf{h}_i}; L(\mathbf{h}_i))$ using $2i(L(\mathbf{h}_i)+3)$ feedback bits. By Proposition 2, we have $|\langle \mathtt{C}_i(\mathbf{h}_i), \mathbf{h}_i \rangle|^2 > \alpha$. This implies that the actual t-dimensional beamforming vector utilized at the transmitter, which is simply constructed by appending $t - i$ zeroes to $\mathtt{C}_i(\mathbf{h}_i)$, will also avoid outage.

By construction, Scheme $\mathtt{D}$ avoids outage for any channel state $\mathbf{h}$ with $\|\mathbf{h}\|^2 > \alpha$. Hence, it achieves the full-CSI outage probability $\mathtt{out}(\mathtt{F})$. Calculations for the training length and feedback rate of Scheme $\mathtt{D}$ are slightly more involved. We present the final results by the following theorem, whose proof can be found in Appendix B.

Theorem 2 *We have* $\mathtt{out}(\mathtt{D}) = \mathtt{out}(\mathtt{F})$ *with* $\mathtt{tl}(\mathtt{D}) \le 1 + \alpha$ *and* $\mathtt{fr}(\mathtt{D}) \le 92(1 + \alpha^3)$.

We shall emphasize that Theorem 2 should be interpreted as "just" an achievability result. Its main message is that the full-CSI performance can be achieved with t-independent training length and feedback rate. Hence, the α-dependent bounds in the statement of Theorem 2 are not necessarily the best-possible as far as a general scheme that can achieve $\mathtt{out}(\mathtt{F})$ is concerned. As can be observed from the proof of the theorem, we have not tried to optimize the bounds.

Let us now also compare the results of Theorem 2 with what we have achieved by Theorem 1 using the antenna selection Scheme B. For Scheme B, we have $\mathtt{tl}(\mathtt{B}), \mathtt{fr}(\mathtt{B}) \in \Theta(e^{\alpha})$ as $\alpha \to \infty$, while for Scheme D, we have $\mathtt{tl}(\mathtt{D}) \in O(\alpha)$ and $\mathtt{fr}(\mathtt{D}) \in O(\alpha^3)$. Hence, there are certain values of t and α where Scheme D improves upon Scheme B in every aspect. It should be clear why Scheme D provides a better outage performance. Regarding the training lengths, note that Scheme B terminates only if the *most-recently* trained antenna avoids outage. On the other hand, Scheme D terminates whenever the joint contribution of *all* trained antennas avoids outage. Therefore, for every channel state, Scheme D always terminates before Scheme B does, and thus, in fact, $\mathtt{tl}(\mathtt{D}) \leq \mathtt{tl}(\mathtt{B})$. The efficiency of Scheme D in terms of training also positively affects its feedback rate: The fewer the amount of antennas that one needs to train, the fewer the feedback messages spent requesting these antennas to be trained. In both cases, same outage probability results in the same diversity.

An interesting special case of Theorem 2 is to assume P is large (but still fixed), and choose $\alpha = P^{m-1}$ for some $m > 1$. Then, if the transmitter has infinitely many antennas (for a simpler discussion, we put the physical impossibility of such an assumption aside), Theorem 2 tells us that we can transmit with rate $\log(1 + P^m) \sim m \log P$ (as $P \to \infty$) outage-free, and thus achieve a multiplexing gain of m. In other words, one can achieve "the MIMO effect" from a MISO system with a very large number of antennas. The price to pay however is a training length of $O(P^m)$ and a feedback rate of $O(P^{3m})$, which are both much larger than the data transmission rate $m \log P$. Ideally, we would like the feedback and training lengths in Theorem 2 (or in another scheme with a $t \to \infty$ vanishing outage probability) to be $o(\log \alpha)$ as $\alpha \to \infty$. Whether this is possible or not will remain as an interesting open problem and shows the need for proving converse results for general interleaved schemes.

On the other hand, regarding the data rate $\log(1 + \alpha P)$, when P is small (a typical case of a low-power system), even slight increase in α significantly improves the data transmission rate. For example, for $P = 1$, increasing α from 1 to 3 doubles the data rate. For such scenarios with

small P, tighter bounds on the training lengths, feedback rates and/or custom-made numerically-designed interleaved schemes are a necessity. In this context, tighter bounds are desirable as they will provide a more accurate estimate on the required training and/or feedback rates to achieve a certain outage probability. On the other hand, numerical designs are desirable as they may outperform the analytically-constructed schemes. Finding an efficient algorithm for the numerical design of interleaved schemes would prove to be a challenging network vector quantization problem [39], where one has to design several interdependent vector quantizers managing the multiple feedback phases of the interleaved scheme. In particular, given t transmitter antennas, one has to design t vector quantizers, $Q_1, \ldots, Q_t$, where the domain of Q_i depends the range of Q_{i-1}. An alternating optimization approach may then be taken where, for the infinite sequence $i = 1, \ldots, t, 1, \ldots, t, \ldots$, one optimizes Q_i while fixing Q_j, $j \neq i$.

V. Quantization Rate Allocation

We now discuss how to further reduce the feedback rate of our proposed schemes using an optimized rate allocation strategy. Recall that in the construction in Section IV-A, one spends a fixed $2(L(\mathbf{h})+3)$ bits per antenna to encode each component of the beamforming vector. Different components of a beamforming vector have different weights in the array gain which is given as $|\langle q(\overrightarrow{\mathbf{h}}; L(\mathbf{h})), \mathbf{h}\rangle|^2$. A component with higher weight should be quantized more accurately, i.e., assigned a higher rate, to provide a better overall performance [40].

For a given beamforming vector $\mathbf{x}$, we assign the optimal quantization rate to each component. To accommodate a variable-rate for different components, we need to adjust the resolution ℓ of the deadzone quantizer. Instead of using the fixed resolution ℓ for all components, resulting in a fixed-rate system, we use the resolution ℓ_{ij} ($i = 1, \cdots, t,\ j = 1, 2$) for the real (if $j = 1$) or imaginary (if $j = 2$) part of x_i. This will result in a variable-rate deadzone quantizer q_v to be defined for an arbitrary beamforming vector $\mathbf{x} = [x_1 \cdots x_t]^T \in \mathbb{C}^t$ with $\|\mathbf{x}\| \leqslant 1$ as $q_v(\mathbf{x}; \boldsymbol{\ell}) \triangleq [\, q(\Re x_1; \ell_{11}) + jq(\Im x_1; \ell_{12}) \ \cdots\ q(\Re x_t; \ell_{t1}) + jq(\Im x_t; \ell_{t2})\,]^T \in \mathbb{C}^t$, where q is the deadzone scalar quantizer and $\boldsymbol{\ell}$ is a $t \times 2$ matrix representing the resolution of q for real and imaginary parts of different components in $\mathbf{x}$. Note that by the definition of the deadzone quantizer q, $|q(x; \ell)| \leq |x|$ for any $x \in [-1, 1]$ and any positive ℓ. Therefore, $\|q_v(\mathbf{x}; \boldsymbol{\ell})\| \leq 1$, which means $q_v(\mathbf{x}, \ell)$ is also a feasible beamforming vector.

Algorithm 1 Rate-Allocation Algorithm

Set $\boldsymbol{\ell}$ to be the $k \times 2$ all-zero matrix, $\overrightarrow{\mathbf{h}}_k = \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$, and $count = 0$.
Set $L(\mathbf{h}_k) = \max\{\lceil \log_2(4k) \rceil, \lceil \log_2 \frac{4k\alpha}{\|\mathbf{h}_k\|^2 - \alpha} \rceil\}$.
while $count < 2k(L(\mathbf{h}_k) + 3)$ **and** $|\langle q_v(\overrightarrow{\mathbf{h}}_k; \boldsymbol{\ell}), \mathbf{h}_k \rangle|^2 < \alpha$ **do**
    $\mathbf{e} = q_v(\overrightarrow{\mathbf{h}}_k; \boldsymbol{\ell} + \boldsymbol{\Delta}) - q_v(\overrightarrow{\mathbf{h}}_k; \boldsymbol{\ell})$, where $\boldsymbol{\Delta} \triangleq \Delta \mathbf{J}_{k\times 2}$, and $\mathbf{J}_{k\times 2}$ is the $k \times 2$ all-one matrix.
    $\mathbf{d}_1 = \Re\overrightarrow{\mathbf{h}}_k \circ \Re\mathbf{e}$, $\mathbf{d}_2 = \Im\overrightarrow{\mathbf{h}}_k \circ \Im\mathbf{e}$.
    Find the indices i and j corresponding to the maximum values of $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively.
    If $\mathbf{d}_1[i] > \mathbf{d}_2[j]$, **then** $\boldsymbol{\ell}[i, 1] = \boldsymbol{\ell}[i, 1] + \Delta$, **else** $\boldsymbol{\ell}[j, 2] = \boldsymbol{\ell}[j, 2] + \Delta$.
    $count = count + \Delta$.
return $q_v(\overrightarrow{\mathbf{h}}_k; \boldsymbol{\ell})$.

To formulate it as a classic rate-allocation problem in a rate-distortion set-up, we define $R_a \triangleq \sum_{i=1}^{k} \sum_{j=1}^{2} \ell_{ij}$, and $D_a \triangleq |\langle q_v(\overrightarrow{\mathbf{h}}_k; \boldsymbol{\ell}), \mathbf{h}_k \rangle|^2 \geqslant \alpha$. The optimal rate-allocation will be achieved by assigning the appropriate quantization rate $\boldsymbol{\ell}$ to each component of $\overrightarrow{\mathbf{h}}_k$ to minimize R_a while satisfying the constraint on D_a. This rate-allocation problem is the dual of the bit-allocation problem in data compression, which is well studied [41]–[43]. Typically, the bit-allocation problem is to minimize the overall distortion under some constraint on the total bit rate while the proposed rate-allocation problem is to minimize the total bit rate under some constraint on the overall distortion. As a result, the generalized Breiman, Friedman, Olshen, and Stone (BFOS) algorithm [43] can be utilized to solve our rate-allocation problem. We design Algorithm 1, based on the generalized BFOS algorithm in [43], to find the optimal rate-allocation to quantize a beamforming vector. The main idea behind the algorithm is as follows. At each step of the algorithm, we assign additional Δ bits to the beamforming vector component that results in the maximum distortion reduction among all possible vector components. This will result in an increase of Δ bits to the total quantization rate and a reduction in the total distortion, i.e., an increase in the array gain. After updating the rate and distortion of the chosen component, we continue the iterations until the overall distortion satisfies the constraint $D_a \geq \alpha$ or the total quantization rate is greater than that of the fixed-rate deadzone quantizer.

VI. Latency Considerations and Antenna Grouping

Our formulations so far ignore the extra latency incurred by dividing the training and feedback stages to multiple stages, as in the proposed interleaved schemes. In this section, we study the latency/performance tradeoffs of interleaving by assuming that every stage of training and interleaving consumes an extra ϵ-fraction of the time that would otherwise be spent on data transmission. This ϵ-cost may, for example, stem from the propagation delays between the transmitter and the receiver during the training and feedback phase.

In such a scenario, training the antennas one by one, as in the previous sections, may be too costly, and thus suboptimal. For this reason, we consider an interleaved antenna selection with antenna grouping that trains antennas K by K, where $K \geq 1$. For simplicity, we assume T is a multiple of K. The transmitter trains the first K antennas and the receiver acquires the CSI for the first K antennas $h_1, \ldots, h_K$. The receiver sends $\lceil \log_2(1+K) \rceil$ bits of feedback that either selects the antenna that can avoid outage or tells the transmitter to train the next K antennas if no such antenna exists. The process continues in the same manner until an antenna that avoids outage is found. If all antennas result in an outage, then the transmitter can simply transmit over an arbitrary antenna. We call this Scheme $\mathtt{B}'$. For the special case of $K = 1$, Scheme $\mathtt{B}'$ is exactly the same as the interleaved antenna selection $\mathtt{B}$ in Section III-B.

Now, suppose that each training/feedback stage costs ϵ-fraction of the channel codeword time. There are totally $\frac{t}{K}$ stages in Scheme $\mathtt{B}'$ so that the channel capacity is $(1 - \frac{t}{K}\epsilon)^+ \log_2(1 + |\langle \mathtt{B}'(\mathbf{h}), \mathbf{h} \rangle|^2 P)$. Given the target data transmission $\rho = \log_2(1+\alpha P)$ as before, the outage probability is given by $\mathrm{Prob}\left(|\langle \mathtt{B}'(\mathbf{h}), \mathbf{h} \rangle|^2 \leq \beta\right)$, where $\beta \triangleq \frac{1}{P}\left((1+\alpha P)^{\frac{1}{(1-\frac{t}{K}\epsilon)^+}} - 1\right)$ can be considered to be a "modified outage threshold" that takes into account cost effects of the training/feedback stages. By the definition of Scheme $\mathtt{B}'$, it follows that an outage occurs if and only if $|h_i|^2 \leq \beta$, $\forall i$, and therefore, we have $\mathtt{out}(\mathtt{B}') = (1 - e^{-\beta})^t$. After some straightforward calculations, we can also obtain the training length and the feedback rate of the scheme

$$\mathtt{tl}(\mathtt{B}') = K\frac{1-(1-e^{-\beta})^t}{1-(1-e^{-\beta})^K}, \qquad \mathtt{fr}(\mathtt{B}') = \lceil \log_2(1+K) \rceil \frac{1-(1-e^{-\beta})^t}{1-(1-e^{-\beta})^K}.$$

in closed form. For the special case of $K = 1$, and β replaced by α, the formulae boil down to the ones provided in Section III-B. Formally analyzing the tradeoffs between $\mathtt{out}(\mathtt{B}')$, $\mathtt{tl}(\mathtt{B}')$, and $\mathtt{fr}(\mathtt{B}')$ for given K and ϵ is not a straightforward task due to the complicated algebraic nature of expressions. Numerical results in the next section, however, suggest that training antennas

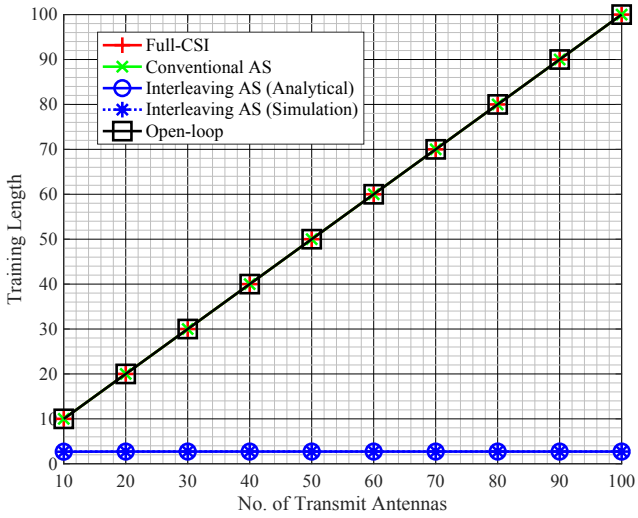

Fig. 6: Training length as a function of t for different schemes in Section III.

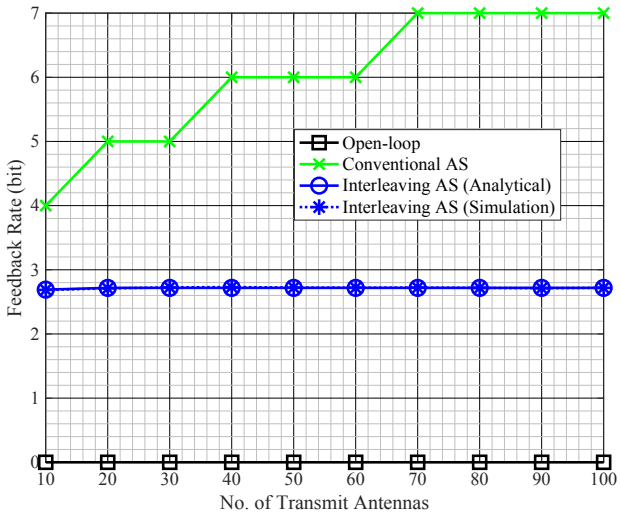

Fig. 7: Feedback rate as a function of t for different schemes in Section III.

one by one is not an optimal strategy in general, and there is an optimal number of antenna groupings K that should be considered.

VII. Simulation Results

In this section, we provide simulation results to compare the performance of different schemes and quantizers. Using rate-allocation results in variable rates for different components of the beamforming vector. We use a Huffman code to send the length of each beamforming vector component. In other words, each resolution, ℓ_{ij}, is Huffman coded and the corresponding prefix-

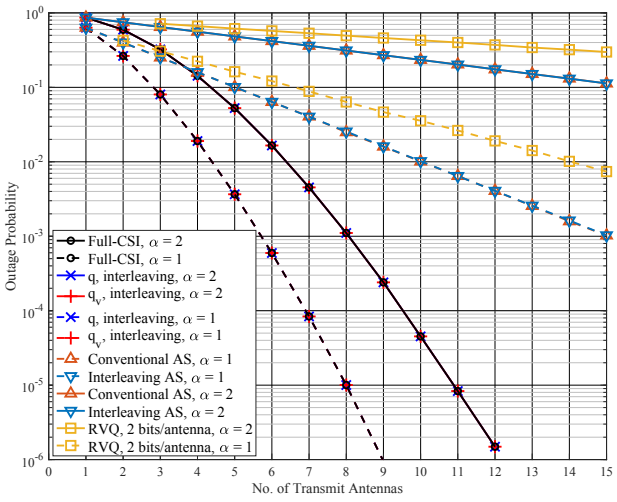

Fig. 8: Outage probability as a function of t for fixed-length and variable-length deadzone quantizers.

free binary codeword representation is sent to the transmitter. In addition, $\Re x_i$ is quantized by $q(\Re x_i; \ell_{i1})$ and $\Im x_i$ is quantized by $q(\Im x_i; \ell_{i2})$, as explained in Section V.

We first present the numerical simulation results of training length and feedback rate as functions of the number of transmit antennas t for different schemes in Section III in Figs. 6 and 7, respectively. We abbreviate antenna selection by AS in both figures. In our simulations, we set $\alpha = 1$. Fig. 6 shows that as t increases, the average training length of the interleaving antenna selection scheme in Section III-B saturates and is lower than those of the full-CSI system, the open-loop system, and the conventional antenna selection scheme in Section III-A. The full-CSI system, the open-loop system, and the conventional antenna selection scheme need to estimate all t channels. Fig. 7 reveals that as t increases, the average feedback rate of the interleaving antenna selection scheme saturates and is lower than those of the full-CSI system and the antenna selection scheme. Note that the feedback rate of the full-CSI system is infinite. For the interleaving antenna selection scheme in both figures, the simulation results align well with the analytical results provided in Theorem 1.

We provide simulation results of the outage probability, the feedback rate, and the average feedback rate as functions of t in Figs. 8, 9, and 10, respectively. We consider the deadzone quantizer $q(\overrightarrow{\mathbf{H}}; \ell)$ and the deadzone quantizer with rate-allocation $q_v(\overrightarrow{\mathbf{H}}; \boldsymbol{\ell})$. The average feedback rate is calculated as the feedback rate divided by the number of transmit antennas. We set $\Delta = 1$ in the rate-allocation algorithm. Fig. 8 demonstrates that the interleaving scheme for both

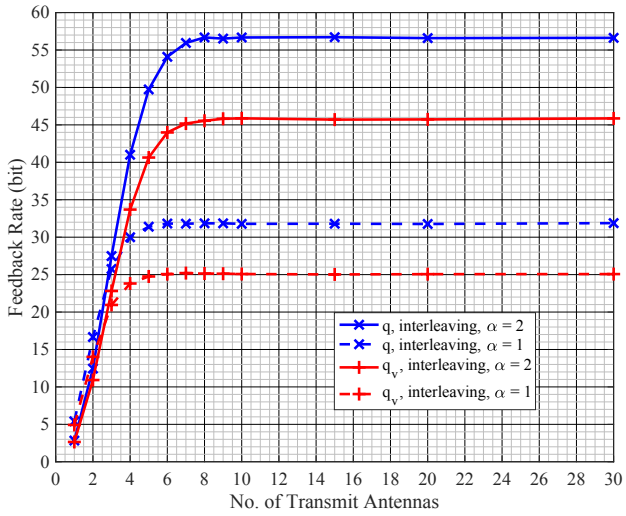

Fig. 9: Feedback rate as a function of t for fixed-length and variable-length deadzone quantizers.

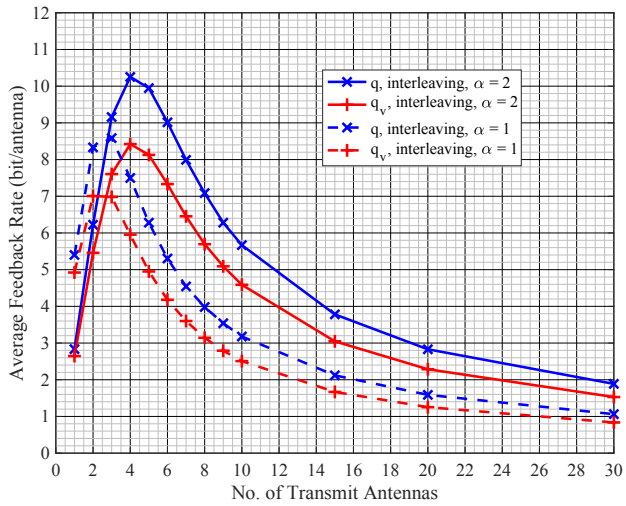

Fig. 10: Average feedback rate as a function of t for fixed-length and variable-length deadzone quantizers.

quantizers can achieve the same outage probability as the full-CSI system. Fig. 8 also shows that the outage probability of the interleaving scheme is better than the outage probabilities of the antenna selection schemes, which is further better than the outage probability of random vector quantization [44] with 2 quantization bits per antenna. A smaller outage threshold α leads to a lower outage probability. Fig. 9 exhibits several important features: First, the feedback rate with interleaving saturates as t increases. Second, the variable-rate deadzone quantizer q_v reduces the total feedback rate compared to the fixed-rate deadzone quantizer q. Third, for the interleaving scheme, the feedback rate decreases as α decreases. This is because a lower resolution for the

beamforming vector is acceptable if the outage threshold decreases. According to Fig. 10, as the number of transmit antennas t increases, the average feedback rate increases when t is small and deceases when t is large. It is shown that the average feedback rates per antenna for both quantizers are approximately equal to or less than 2 bits/antenna when t is large.

According to Figs. 9 and 10, the feedback rates of both deadzone quantizers saturate as the number of transmit antennas increases. This is a key difference compared to the conventional CSI quantization techniques for massive MIMO systems. For example, using the method proposed in [19], the receiver sends back a binary feedback sequence of length $Bt + q$ where B is the number of quantization bits used per transmit antenna and q is a small positive constant, which scales linearly with the number of transmit antennas. As a result, compared to the conventional CSI quantizers, the proposed deadzone quantizers can save a large amount of feedback overhead when the number of transmit antennas is large.

For Scheme $\mathtt{B}'$ of Section VI, we present the outage probability, the training length, and the feedback rate as functions of the number of trained antennas at a time, K, in Figs. 11, 12, and 13, respectively. We can observe that the analytical results match with the simulations in all cases. In Fig. 11, the outage probability decreases with K since the SNR threshold β is a decreasing function of K, and $\mathtt{out}(\mathtt{B}')$ decreases as β decreases. As expected, as the per-stage cost ϵ increases, the outage probability increases. Also, according to Fig. 12, as K increases from 1 to 30, the training length decreases at first but then increases. The optimal value of K that minimizes the training length is 2 for $\epsilon = 0.01$ and 3 for $\epsilon = 0.02$. According to Fig. 13, the optimal value of K that minimizes the feedback rate is 3 for $\epsilon = 0.01$ and 6 for $\epsilon = 0.02$. According to these results, it is suboptimal to train the antennas one by one for the particular choices of the system parameters in Figs. 11, 12, and 13. Depending on design requirements, one should consider grouping the antennas in the training and feedback phases.

VIII. Conclusion

We introduced and analyzed multi-antenna communication schemes whose training and feedback stages are interleaved and mutually interacting. We applied the interleaving scheme to MISO systems to achieve the same outage probability as the full-CSI system using partial CSIT and partial CSIR. We designed a deadzone quantizer and a rate-allocation algorithm to send the feedback messages by a limited number of feedback bits. With t transmit antennas, the

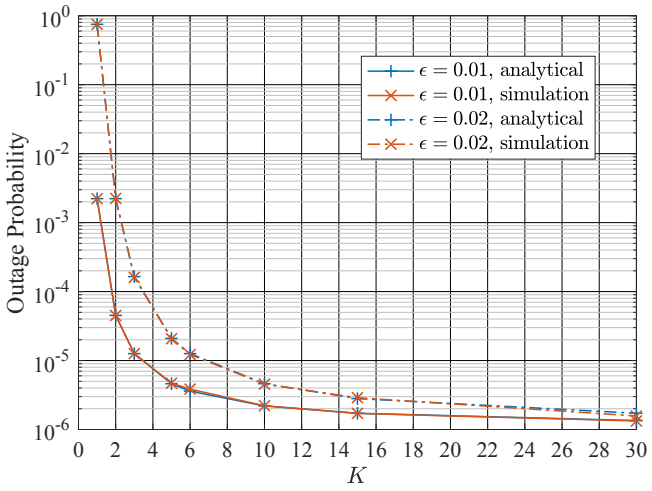

Fig. 11: Outage probability as a function of K for Scheme B′ when $t = 30$, $\alpha = 1$, $P = 1$, $\epsilon = 0.01$ or 0.02.

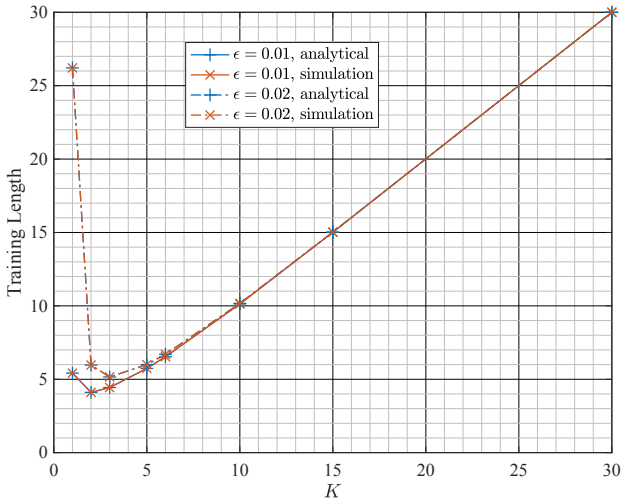

Fig. 12: Training length as a function of K for Scheme B′ when $t = 30$, $\alpha = 1$, $P = 1$, $\epsilon = 0.01$ or 0.02.

interleaving scheme with the deadzone quantizer can achieve a t-independent finite feedback rate which only depends on the power constraint and the target data rate. In addition, the rate-allocation algorithm can further reduce the feedback rate by assigning distinct quantization rates to different components in a beamforming vector.

The idea of interleaving can also be used in conjunction with rate adaptation. Suppose the rate-adaptive system can support a number of rates, say, $\rho_1, \ldots, \rho_n$, that one can choose from. Receiver feedback will then be used to choose the beamforming vector as well as the transmission rate. An outage can be declared if the system cannot even support the minimum $\min_{i\in\{1,\ldots,n\}} \rho_i$ of data rates. Given a certain outage probability, one can then study the tradeoff between the

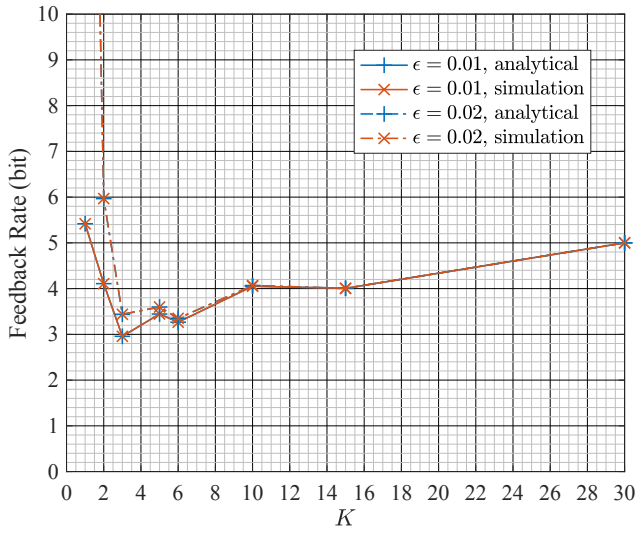

Fig. 13: Feedback rate as a function of K for Scheme B′ when $t = 30$, $\alpha = 1$, $P = 1$, $\epsilon = 0.01$ or 0.02.

feedback rates, training lengths, and the average data transmission rate. For example, supporting high rates typically requires more CSI, and thus larger feedback rates and/or training lengths.

Also, in this work, we have only considered a total power constraint across all antennas. The performance of interleaved training and limited feedback schemes with the additional per-antenna power constraints is another direction for future research. Another interesting topic is the design and analysis of interleaved beam selection schemes for multi-carrier systems such as OFDM.

APPENDIX A
PROOF OF PROPOSITION 1

We first determine the $t \to \infty$ asymptotic behavior of $\mathtt{out}(\mathtt{F})$. For this purpose, note that

$$\mathtt{out}(\mathtt{F}) = \mathrm{P}(\|\mathbf{h}\|^2 \leq \alpha) = \sum_{i=t}^{\infty} \frac{\alpha^i e^{-\alpha}}{i!}, \tag{3}$$

which leads to an easy lower bound (by considering only the $i = t$ term) $\mathtt{out}(\mathtt{F}) \geq \frac{\alpha^t e^{-\alpha}}{t!}$. For an upper bound, we can rewrite (3) as

$$\mathtt{out}(\mathtt{F}) = \frac{\alpha^t e^{-\alpha}}{t!} \sum_{i=0}^{\infty} \frac{\alpha^i}{(t+1)\cdots(t+i)}. \tag{4}$$

Since $(t+1)\cdots(t+i) \geq i!$, we obtain $\mathrm{P}(\|\mathbf{h}\|^2 \leq \alpha) \leq \frac{\alpha^t}{t!}$. Combining the upper and lower bounds, we have $\mathtt{out}(\mathtt{F}) \in \Theta(\frac{\alpha^t}{t!})$, as desired.

We now determine the outage probability of an open-loop system as $t \to \infty$. We recall that $\mathtt{out}(\mathtt{G}) = \mathrm{P}(\|\mathbf{h}_{\kappa(t)}\|^2 \leq \kappa(t)\alpha)$, where $\kappa(t) \triangleq \arg\min_{k\in\{1,\ldots,t\}} \mathrm{P}(\|\mathbf{h}_k\|^2 \leq k\alpha)$ with ties broken

in favor of k with the smallest index. Then, either $\kappa(t) = t$ for infinitely many t or $\exists t_0 \geq 1$, $\forall t \geq t_0$, $\kappa(t) = t_0$. For values of α that satisfy the latter scenario, we have $\mathtt{out}(\mathtt{G}) = \Theta(1)$.

Suppose $0 < \alpha < 1$. It follows from (3) that $\mathrm{P}(\|\mathbf{h}\|^2 \leq t\alpha) \geq \frac{(t\alpha)^t e^{-t\alpha}}{t!}$. On the other hand, substituting $t\alpha$ instead of α to the expansion in (4), and using the bound $(t+1)\cdots(t+i) \geq t^i$ for the denominator of the fraction in summation, we obtain $\mathrm{P}(\|\mathbf{h}\|^2 \leq t\alpha) \leq \frac{(t\alpha)^t e^{-t\alpha}}{t!(1-\alpha)}$. Combining the upper and lower bounds, it follows that we have $\mathtt{out}(\mathtt{G}) = \Theta(\frac{(t\alpha)^t e^{-t\alpha}}{t!})$ for $0 < \alpha < 1$.

Now, suppose $\alpha \geq 1$. In this case, the Berry-Esseen theorem provides the estimate $|P(\|\mathbf{h}\|^2 \leq t\alpha) - \Phi((\alpha-1)\sqrt{t})| \leq \frac{C}{\sqrt{t}}$ for some constant $C > 0$, where $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution with mean 0 and variance 1. It follows that $P(\|\mathbf{h}\|^2 \leq t) \to \frac{1}{2}$ when $\alpha = 1$, and $P(\|\mathbf{h}\|^2 \leq t\alpha) \to 1$ whenever $\alpha > 1$. Hence, $\mathtt{out}(\mathtt{G}) = \Theta(1)$ for $\alpha \geq 1$.

APPENDIX B
PROOF OF THEOREM 2

The fact that $\mathtt{out}(\mathtt{D}) = \mathtt{out}(\mathtt{F})$ follows immediately. We thus first calculate the training length $\mathtt{tl}(\mathtt{D})$ of Scheme $\mathtt{D}$. Let $\mathcal{A}_1 \triangleq \{\mathbf{h} \in \mathbb{C}^t : |h_1|^2 > \alpha\}$, $\mathcal{A}_i \triangleq \{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}_i\|^2 > \alpha,\ \|\mathbf{h}_{i-1}\|^2 \leq \alpha\}$, $i = 2, \ldots, t$, and $\mathcal{B} \triangleq \{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}_t\|^2 \leq \alpha\}$. Note that the sets $\mathcal{A}_1, \ldots, \mathcal{A}_t, \mathcal{B}$ form a partition of $\mathbb{C}^t$. For any $i \in \{1, \ldots, t\}$, if $\mathbf{h} \in \mathcal{A}_i$, the transmitter trains only the first i channels $\mathbf{h}_1, \ldots, \mathbf{h}_i$. If $\mathbf{h} \in \mathcal{B}$, the transmitter trains all the t channels. The training length is thus

$$\mathtt{tl}(\mathtt{D}) = \sum_{i=1}^{t} i\mathrm{P}(\mathbf{h} \in \mathcal{A}_i) + t\mathrm{P}(\mathbf{h} \in \mathcal{B}), \tag{5}$$

We have $\mathrm{P}(\mathbf{h} \in \mathcal{A}_1) = e^{-\alpha}$. For $i \in \{2, \ldots, t-1\}$, we have

$$\mathrm{P}(\mathbf{h} \in \mathcal{A}_i) = \int_0^{\alpha} \int_{\alpha-x}^{\infty} e^{-y} \frac{e^{-x} x^{i-2}}{(i-2)!} \mathrm{d}y \mathrm{d}x = e^{-\alpha} \int_0^{\alpha} \frac{x^{i-2}}{(i-2)!} \mathrm{d}x = \frac{\alpha^{i-1} e^{-\alpha}}{(i-1)!}. \tag{6}$$

Also, since $t\mathrm{P}(\mathbf{h} \in \mathcal{B}) = \sum_{i=t}^{\infty} t\frac{\alpha^i e^{-\alpha}}{i!} \leq \sum_{i=t}^{\infty} \frac{\alpha^i e^{-\alpha}}{(i-1)!} = \alpha \sum_{i=t-1}^{\infty} \frac{\alpha^i e^{-\alpha}}{i!}$, we have

$$\begin{aligned}
\mathtt{tl}(\mathtt{D}) &\leq e^{-\alpha} \left(\sum_{i=1}^{t} \underbrace{i}_{=(i-1)+1} \frac{\alpha^{i-1}}{(i-1)!} + \alpha \sum_{i=t-1}^{\infty} \frac{\alpha^i}{i!} \right) \\
&= e^{-\alpha} \left(\sum_{i=2}^{t} \frac{\alpha^{i-1}}{(i-2)!} + \sum_{i=1}^{t} \frac{\alpha^{i-1}}{(i-1)!} + \alpha \sum_{i=t-1}^{\infty} \frac{\alpha^i}{i!} \right) = e^{-\alpha} \Bigg(\alpha \underbrace{\sum_{i=0}^{\infty} \frac{\alpha^i}{i!}}_{=e^{\alpha}} + \underbrace{\sum_{i=1}^{t} \frac{\alpha^{i-1}}{(i-1)!}}_{\leq e^{\alpha}} \Bigg) \leq 1 + \alpha,
\end{aligned}$$

as claimed in the statement of the theorem.

We now calculate the feedback rate $\mathtt{fr}(\mathtt{D})$ of Scheme D. Note that for any $i = 1, \ldots, t$, if $\mathbf{h} \in \mathcal{A}_i$, the receiver sends a total of $(i-1)$ bits for requesting the transmitter to train the first $i-1$ antennas (via $i-1$ one-bit binary codewords "0"). In addition, it sends $2i(L(\mathbf{h}_i)+3)$ bits for the outage avoiding quantized beamforming vector, for a total of $2iL(\mathbf{h}_i)+7i-1$ feedback bits. For $\mathbf{h} \in \mathcal{B}$, there are only t feedback bits. The feedback rate is thus given by

$$\begin{aligned}\mathtt{fr}(\mathtt{D}) &= \sum_{i=1}^{t} \int_{\mathcal{A}_i} (2iL(\mathbf{h}_i)+7i-1) f(\mathbf{h}) \mathrm{d}\mathbf{h} + t\mathrm{P}(\mathbf{h} \in \mathcal{B}) \\ &= 2\sum_{i=1}^{t} \int_{\mathcal{A}_i} iL(\mathbf{h}_i) f(\mathbf{h}) \mathrm{d}\mathbf{h} + \sum_{i=1}^{t} (7i-1)\mathrm{P}(\mathbf{h} \in \mathcal{A}_i) + t\mathrm{P}(\mathbf{h} \in \mathcal{B}),\end{aligned}$$

where $f(\mathbf{h})$ represents the probability density function of $\mathbf{h}$. According to (5), the sum of the last two terms can be upper bounded by $7\mathtt{tl}(\mathtt{D}) \le 7(1+\alpha)$. Therefore,

$$\mathtt{fr}(\mathtt{D}) \le 7(1+\alpha) + 2\sum_{i=1}^{t} \int_{\mathcal{A}_i} iL(\mathbf{h}_i) f(\mathbf{h}) \mathrm{d}\mathbf{h}.$$

We now evaluate the sum. For this purpose, we partition $\mathcal{A}_1, \ldots, \mathcal{A}_t$ via $\mathcal{A}_i' \triangleq \{\mathbf{h} \in \mathbb{C}^t : \alpha < \|\mathbf{h}_i\|^2 < 2\alpha,\ \|\mathbf{h}_{i-1}\|^2 \le \alpha\}$ and $\mathcal{A}_i'' \triangleq \{\mathbf{h} \in \mathbb{C}^t : \|\mathbf{h}_i\|^2 \ge 2\alpha,\ \|\mathbf{h}_{i-1}\|^2 \le \alpha\}$, with the convention that $\mathbf{h}_0 = 0$ is deterministic. Note that for any $i \in \{1, \ldots, t\}$, if $\mathbf{h} \in \mathcal{A}_i'$, then

$$\begin{aligned}L(\mathbf{h}_i) = \left\lceil \log_2 \frac{4i\alpha}{\|\mathbf{h}_i\|^2 - \alpha} \right\rceil &\le 1 + \log_2 \frac{4i\alpha}{\|\mathbf{h}_i\|^2 - \alpha} \\ &= 3 + \underbrace{\frac{1}{\log 2}}_{\le 2} \log \frac{i\alpha}{\|\mathbf{h}_i\|^2 - \alpha} \le 3 + 2\log i + 2\log \frac{\alpha}{\|\mathbf{h}_i\|^2 - \alpha}, \quad (7)\end{aligned}$$

while if $\mathbf{h} \in \mathcal{A}_i''$, then $L(\mathbf{h}_i) = \lceil \log_2(4i) \rceil \le 3 + 2\log i$. Thus,

$$\mathtt{fr}(\mathtt{D}) \le 7(1+\alpha) + 6\underbrace{\sum_{i=1}^{t} i\mathrm{P}(\mathbf{h} \in \mathcal{A}_i)}_{\triangleq S_1} + 4\underbrace{\sum_{i=1}^{t} i \log i \mathrm{P}(\mathbf{h} \in \mathcal{A}_i)}_{\triangleq S_2} + 4\sum_{i=1}^{t} \underbrace{\int_{\mathcal{A}_i'} i \log \frac{\alpha}{\|\mathbf{h}_i\|^2 - \alpha} f(\mathbf{h}) \mathrm{d}\mathbf{h}}_{\triangleq S_{3i}}. \quad (8)$$

We now find upper bounds on S_1, S_2, and $\sum_{i=1}^{t} S_{3i}$. Regarding S_1 and S_2, note that we have already evaluated the probabilities $\mathrm{P}(\mathbf{h} \in \mathcal{A}_i)$, $i = 1, \ldots, t$ in (6). Hence,

$$S_1 = \underbrace{e^{-\alpha}}_{\le 1} + \alpha e^{-\alpha} \underbrace{\sum_{i=2}^{t}}_{\le \sum_{i=2}^{\infty}} \underbrace{\frac{i}{i-1}}_{\le 2} \frac{\alpha^{i-2}}{(i-2)!} \le 1 + 2\alpha, \quad (9)$$

$$S_2 = e^{-\alpha} \sum_{i=2}^{t} i \log i \frac{\alpha^{i-1}}{(i-1)!} = e^{-\alpha} \underbrace{\sum_{i=2}^{t}}_{\le \sum_{i=2}^{\infty}} \underbrace{\frac{i}{i-1}}_{\le 2} \log i \frac{\alpha^{i-1}}{(i-2)!} \le 2\alpha e^{-\alpha} \sum_{i=0}^{\infty} \log(i+2) \frac{\alpha^i}{i!}$$

$$= \frac{2\alpha}{e^{\alpha}} \left(\sum_{i=0}^{\lceil \alpha \rceil} \underbrace{\log(i+2)}_{\leq \log(\lceil \alpha \rceil + 2)} \frac{\alpha^i}{i!} + \sum_{i=\lceil \alpha \rceil + 1}^{\infty} \underbrace{\frac{\log(i+2)}{i}}_{\leq \frac{\log(\lceil \alpha \rceil + 2)}{\lceil \alpha \rceil}} \frac{\alpha^i}{(i-1)!} \right)$$

$$\leq 2\alpha e^{-\alpha} \underbrace{\log(\lceil \alpha \rceil + 2)}_{\leq \log(\alpha+3)} \left(\sum_{i=0}^{\lceil \alpha \rceil} \frac{\alpha^i}{i!} + \underbrace{\frac{\alpha}{\lceil \alpha \rceil}}_{\leq 1} \sum_{i=\lceil \alpha \rceil + 1}^{\infty} \frac{\alpha^{i-1}}{(i-1)!} \right)$$

$$\leq 2\alpha e^{-\alpha} \log(\alpha+3) \left(\underbrace{\sum_{i=0}^{\lceil \alpha \rceil} \frac{\alpha^i}{i!}}_{\leq e^{\alpha}} + \underbrace{\sum_{i=\lceil \alpha \rceil + 1}^{\infty} \frac{\alpha^{i-1}}{(i-1)!}}_{\leq e^{\alpha}} \right) \leq 4\alpha \log(\alpha+3). \quad (10)$$

For an upper bound on $\sum_{i=1}^{t} S_{3i}$, we consider S_{31}, S_{32}, and $\sum_{i=3}^{t} S_{3i}$ separately. We have

$$S_{31} = \int_{\alpha}^{2\alpha} \log \frac{\alpha}{x-\alpha} \underbrace{e^{-x}}_{\leq 1} \mathrm{d}x \leq \int_{\alpha}^{2\alpha} \log \frac{\alpha}{x-\alpha} \mathrm{d}x = \alpha, \quad (11)$$

$$S_{32} = 2 \int_{0}^{\alpha} \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} \underbrace{e^{-y} e^{-x}}_{\leq 1} \mathrm{d}y \mathrm{d}x \leq 2 \int_{0}^{\alpha} \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} \mathrm{d}y \mathrm{d}x = 2\alpha^2, \quad (12)$$

$$\sum_{i=3}^{t} S_{3i} = \sum_{i=3}^{t} i \int_{0}^{\alpha} \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} e^{-y} \frac{x^{i-2} e^{-x}}{(i-2)!} \mathrm{d}y \mathrm{d}x$$

$$= \int_{0}^{\alpha} x \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} e^{-y} \underbrace{\sum_{i=3}^{t}}_{\leq \sum_{i=3}^{\infty}} \underbrace{\frac{i}{i-2}}_{\leq 3} \frac{x^{i-3} e^{-x}}{(i-3)!} \mathrm{d}y \mathrm{d}x$$

$$\leq 3 \int_{0}^{\alpha} x \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} \underbrace{e^{-y}}_{\leq 1} \underbrace{\sum_{i=3}^{\infty} \frac{x^{i-3} e^{-x}}{(i-3)!}}_{=1} \mathrm{d}y \mathrm{d}x$$

$$\leq 3 \int_{0}^{\alpha} x \int_{\alpha-x}^{2\alpha-x} \log \frac{\alpha}{x+y-\alpha} \mathrm{d}y \mathrm{d}x = 3\alpha \int_{0}^{\alpha} x \mathrm{d}x = \frac{3}{2} \alpha^3. \quad (13)$$

Substituting the bounds in (9), (10), (11), (12), and (13) to (8), we obtain $\mathtt{fr}(\mathtt{D}) \leq 13 + 23\alpha + 16\alpha \log(3+\alpha) + 8\alpha^2 + 6\alpha^3$. Using the bound $\log(3+\alpha) \leq \alpha + 2$, we obtain $\mathtt{fr}(\mathtt{D}) \leq 13 + 55\alpha + 24\alpha^2 + 6\alpha^3$. Finally, the inequalities $\alpha^2 \leq 1 + \alpha^3$ and $\alpha \leq 1 + \alpha^3$ lead to $\mathtt{fr}(\mathtt{D}) \leq 92 + 85\alpha^3 \leq 92(1 + \alpha^3)$, and this concludes the proof.

REFERENCES

[1] E. Koyuncu and H. Jafarkhani, "Interleaving training and limited feedback for point-to-point massive multiple antenna systems," *IEEE Intl. Symp. Inf. Theory*, June 2015.

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

[3] 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.

[4] 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.

[5] 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.

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

[7] L. Liu and H. Jafarkhani, “Application of quasi-orthogonal space-time block codes in beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 54–63, Jan. 2005.

[8] 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.

[9] 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.

[10] L. P. Natarajan and B. S. Rajan, “Full-rate full-diversity finite feedback space-time schemes with minimum feedback and transmission duration,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5022–5034, Oct. 2013.

[11] R. Rajashekar, Marco Di Renzo, K. V. S. Hari, and L. Hanzo, “A generalized transmit and receive diversity condition for feedback-assisted MIMO systems: theory and applications in full-duplex spatial modulation,” *IEEE Trans. Signal Process.*, vol. 65, no. 24, pp. 6505–6519, Dec.15, 2017.

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

[13] 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.

[14] 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.

[15] M. R. A. Khandaker and Y. Rong, “Precoding design for MIMO relay multicasting,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544–3555, July 2013.

[16] E. Koyuncu, C. Remling, X. Liu, and H. Jafarkhani, “Outage-optimized multicast beamforming with distributed limited feedback,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2069–2082, Apr. 2017.

[17] J. Hoydis, S. T. Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?” *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[18] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[19] J. Choi, Z. Chance, D. J. Love, and U. Madhow, “Noncoherent trellis coded quantization: a practical limited feedback technique for massive MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[20] X. Rao and V. K. N. Lau, “Distributed compressive CSIT estimation and feedback for FDD multi-User massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, June 2014.

[21] P.-H. Kuo, H. T. Kung, and P.-A. Ting, “Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays,” *Proc. IEEE WCNC*, Shanghai, China, Apr. 2012, pp. 1–6.

[22] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, Jr., “Compressed channel feedback for correlated massive MIMO systems,” *J. Commun. Net.*, vol. 18, no. 1, pp. 95–104, 2016.

[23] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: open-Loop and closed-Loop training with memory," *IEEE J. Select. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[24] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: realizing massive MIMO gains with limited channel state information," *Proc. IEEE CISS*, Pacific Grove, CA, USA, pp. 1–6, Nov. 2013.

[25] B. Lee, J. Choi, J.-Y. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," vol. 63, no. 9, pp. 3261–3274, Sept. 2015.

[26] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[27] Y. Ren, Y. Wang, C. Qi, and Y. Liu, "Multi-beam selection with limited feedback for hybrid beamforming in massive MIMO systems", *IEEE Access*, vol. 5, pp. 13327–13335, Feb. 2017.

[28] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, Jan. 2016.

[29] L. Liu and H. Jafarkhani, "Transmit beamforming for a large reconfigurable antenna array," *IEEE Global Commun. Conf.*, Nov. 2005.

[30] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.

[31] U. Salim, D. Gesbert, and D. Slock, "Combining training and quantized feedback for multiantenna reciprocal channels," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1383–1396, Mar. 2012.

[32] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Joint signal and channel state information compression for the backhaul of uplink network MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1555–1567, Jan. 2014.

[33] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Interleaved training and training-based transmission design for hybrid massive antenna downlink," *IEEE J. Select. Topics Signal Process.*, vol. 12, no. 3, pp. 541–556, June 2018.

[34] 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.

[35] E. Koyuncu and H. Jafarkhani, "Delay-limited and ergodic capacities of MIMO channels with limited feedback," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3682–3696, July 2016.

[36] H. Jafarkhani, Space-Time Coding: Theory and Practice, *Cambridge University Press*, 2005.

[37] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. on Commun.*, vol. 63, no. 11, pp. 3917-3928, July 2015.

[38] E. Abbe, S.-L. Huang, and E. Telatar, "Proof of the outage probability conjecture for MISO channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2596–2602, Apr. 2013.

[39] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1584–1604, Aug. 2004.

[40] R. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[41] P. H. Westerink, J. Biemond, and D.E. Boekee, "An optimal bit allocation algorithm for sub-band coding," *Proc. IEEE ICASSP*, New York, NY, USA, Apr. 1988, pp. 1–6.

[42] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers (speech coding)," *IEEE Trans. Acoustics, Speech, Signal Procss.*, vol. 36, no. 9, pp. 1445–1453, Sept. 1988.

[43] E. A. Riskin, "Optimal bit allocation via the generalized BFOS algorithm," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 400-402, Mar. 1991.

[44] W. Santipach and M. L. Honig, "Capacity of multiple-antenna fading channel with quantized precoding matrix," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1218–1234, Mar. 2009.