Outage Theorems for MIMO Block Fading Channels

Narayan Prasad and Mahesh K. Varanasi

August 9, 2006

Abstract

The connection between the average codeword or frame error probability (FEP) of space-time codes and the outage probability over general block fading multi-input multi-output (MIMO) channels is established. Three archetypal problems are considered under general fading distributions in a single framework wherein the receiver has channel state information whereas the transmitter knows (a) the fading distribution but not the channel realization (b) the channel realization but must follow a short term (per codeword) average power constraint and (c) the channel realization but is constrained only by a long-term average power constraint. Three telescoping sets of space-time codes are defined for a given rate and it is shown that average FEPs arbitrarily close to the respective outage probabilities for each of the three cases (a)-(c) can be achieved by codes in each set for sufficiently large framelengths. For the smallest set among the three which contains codes with a spectral norm constraint that is stricter than the average or maximum energy constraints commonly assumed, firm sphere-packing lower bounds on the FEP are obtained, and consequently, strong converse theorems are proved which assert that the respective outage probabilities also represent the *best achievable* FEP in the large framelength limit. Moreover, the set of spectral norm constrained codes are also shown to be large enough to contain *universal codes* that can communicate reliably over any channel realization for which the mutual information exceeds the information rate of the code.

Index Terms: Frame error probability, MIMO, MIMO-OFDM, multiple antenna channels, outage probability, fading channels, space-time codes, universal codes.

This work was supported in part by NSF Grant CCF-0431170. The authors are with the Electrical and Computer Engineering Department, University of Colorado, Boulder, CO 80309 USA (email:{prasadn, varanasi}@dsp.colorado.edu).

I. INTRODUCTION

Dealing with fading is one of the primary challenges in wireless communication systems. In many practical cases where the channel characteristics change slowly, the transmission of each codeword spans only a limited number of fading realizations and is commonly modeled by an L-block fading channel. The results of this paper are applicable for all such channels in which it is not possible to guarantee a given nonzero transmission rate with an arbitrarily small probability of error. These include all channels whose delay-limited capacity[1] [1, 2] is zero, or more generally, channels over which the rate of transmission employed is higher than the delay-limited capacity. In such channels, it is the outage probability that is the primary measure of interest. The outage probability is defined as the infimum of the probability that the instantaneous mutual information of the channel falls below the transmission rate. Systems for which the delay-limited capacity is positive and the rate selected is less than the delay-limited capacity, the outage probability would be zero and proving the strong converse theorem would be trivial since the error probability of any code is non-negative and hence lower bounded by the outage probability. Thus, without loss of generality, we restrict our attention to systems whose outage probability for the given rate and power is strictly positive [2].

It is well understood that the outage probability, denoted by $P_{\text{out}}(R)$, is achievable in the sense that for any $\epsilon > 0$, there exists a code of sufficiently large block or frame length for which the average frame error probability is upper bounded by $P_{\text{out}}(R) + \epsilon$. In many works, see for instance [3, 4], outage probability is also stated to be the *best achievable* probability of error in the limit of large codeword length, which implicitly assumes that a strong converse holds true. While such a strong converse was proved by the authors in [5] for single-input multi-output (SIMO) channels under the maximum energy constraint using classical results from [6, 7], to the best of our knowledge, no such strong converse for MIMO channels has been proved under either the average or the maximum energy constraint. Thus even though outage probability for MIMO systems has been widely accepted as an indicator of the optimal frame error probability

---

[1] The delay-limited capacity (also known as the non-ergodic capacity or the min-capacity) is defined to be the maximum rate such that the resulting outage probability (defined appropriately for the given channel state information availability and given power constraints, short- or long-term) is zero.

[2] The delay-limited capacity is well defined in that it has a coding theorem and a converse, and is zero for all the commonly used fading channel models in the literature in the absence of channel state information at the transmitter and a long-term power constraint [1, 2].

in the limit of large codeword length, only the achievability has been rigorously proved, from which one can only conclude that the outage probability is an upper bound on the best-achievable frame error rate. The more desirable result would be the strong converse that involves deriving a lower bound which together with the achievability result would mean that not only can the frame error rate reach the outage probability but also that it cannot be lower and hence that the outage probability is a *fundamental limit*.

Recall that the usefulness of the restricted class of codes satisfying the maximum energy constraint (as opposed to the average energy constraint) for Gaussian single-input, single-output (SISO) channels arises from the twin facts that the strong converse theorem can be proved for this class *and* that this class is large enough to prove the achievability of any rate less than the channel capacity. In the same way, we define a somewhat more restricted class of codes (with a spectral norm constraint and expanded via precoding) than even those that satisfy the maximum energy constraint for slow-fading MIMO channels but for which we are able to establish both the achievability and strong converse theorems. For the special case of SIMO channels, our class of codes is identical to that satisfying the maximum energy constraint, but is strictly smaller for the MIMO channel.

Our class of codes with the spectral norm constraint is large enough however, to encompass any type of precoding that includes beamforming and/or power control. Moreover, we deal with MIMO channels with general fading distributions and with problems associated with channel state information only at the receiver (CSIR) [8] as well as with channel state information at the transmitter (CSIT) under either the short-term or long-term energy constraints [2].

Since our interest here is on delay-limited systems where each codeword sees one or finitely many channel realizations, the capacity of such systems is best explained in the capacity versus outage formulation where the block fading channel is modeled as a compound channel characterized by the set of all possible fading realizations. The notion of universality is meaningful in the compound channel scenario, where we say a code is universal over a set of channel realizations if it results in an acceptably small (conditional) error probability *uniformly* over that set. The seminal work of [9] shows the existence of codes of rate R which are universal over any bounded set of channel realizations for which the delay-limited capacity is no less than R. This achievability result is of course stronger than the one obtained using random coding arguments which yields the existence of a code whose *average* error probability (averaged over the set) is acceptably small.

Strengthening the achievability result of [9] which was proved for the set of maximum energy constrained codes, we show that even the smaller class of spectral norm constrained codes contain *universal codes*. The practical significance of this result and the one obtained in [9] as explained in Section IV is a kind of "separation principle" that states it is possible to *decouple* the design of optimal codes (in the sense of achieving average FEPs arbitrarily close to the respective outage probabilities) into the design of spectral norm or maximum energy constrained universal codes and the design of precoders optimized for the particular fading distribution and the availability or the lack thereof, of the channel state information at the transmitter.

The following notation is used throughout this paper. Vectors and matrices are represented by boldface lower and upper case letters, respectively. The super-scripts $(\cdot)^T$ and $(\cdot)^\dagger$ denote the transpose and the conjugate transpose operations, respectively. For any $n, p \geq 1$, $\mathbb{C}^{n\times p}$, denotes the set of $n \times p$ matrices with complex-valued components. $\mathbf{1}_p$ and $\mathbf{0}_p$ are used to denote length p vectors of ones and zeros, respectively. We let $\mathrm{tr}(\cdot)$ and $|\cdot|$ denote the trace and determinant of their matrix arguments whereas $E[\cdot]$ denotes the expectation operator. For $\mathbf{A} \in \mathbb{C}^{n\times p}$, $\mathbf{a} \in \mathbb{C}^{n\times 1}$, we define $\|\mathbf{A}\|_{\mathrm{F}}^2 \overset{\triangle}{=} \mathrm{tr}(\mathbf{A}^\dagger\mathbf{A})$ and $\|\mathbf{a}\|^2 \overset{\triangle}{=} \mathbf{a}^\dagger\mathbf{a}$. Also, $\|\cdot\|_\infty$ and $\lambda_{\max}(\cdot)$ denote the maximum singular value and the maximum eigenvalue of their respective matrix arguments. Finally, $\log(\cdot)$ denotes the logarithm of its argument to base 2.

II. SYSTEM MODEL

We consider the discrete-time model of a block-fading MIMO communication system. In particular, the system has K inputs, N outputs and the channel output received over J channel uses can be described as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}\,, \tag{1}$$

where $\mathbf{Y}$ is the $N \times J$ received matrix and the fading is described by the $N \times K$ matrix $\mathbf{H}$. We assume that the random variables (fading coefficients) in the matrix $\mathbf{H}$ are drawn from some continuous distribution and remain fixed for J symbol intervals after which they jump to independent values. Due to delay and/or bandwidth constraints each transmitted codeword sees only finitely many fading realizations. Note that in (1) we can have $N = Lr$ and $K = Lt$, where

r, t denote the number of receive and transmit antennas, respectively, and L denotes the finite number of fading realizations. For example, L may denote the number of carriers in MIMO orthogonal frequency-division multiplexed (OFDM) systems. The matrix $\mathbf{H}$ in this case has a block diagonal structure and can be considered to be the channel matrix of a K-input and N-output MIMO system with a coherence interval of J symbol intervals. The $N \times J$ matrix $\mathbf{V}$ represents the additive noise at the receiver which is independent of $\mathbf{H}$ and has i.i.d. complex normal $\mathcal{CN}(0, 1)$ elements. The instantaneous channel state information (CSI) is assumed to be available perfectly at the receiver (CSIR) but may or may not be known to the transmitter. The transmitter however always knows the distribution (or the fading law) of $\mathbf{H}$. The $K \times J$ codeword matrix $\mathbf{X}$ is drawn equi-probably from a space-time code $\mathcal{X}_J$ and the decoder used is the optimum decoder employing the maximum-likelihood decoding rule.

Our objective here is to examine the connection between the outage probability and the FEP in the limit of large block-lengths $J \to \infty$. However, since J is proportional to the coherence time of the underlying physical channel, it does not make sense to conduct the performance analysis for a *given channel* by letting $J \to \infty$. The correct approach—which is adopted here henceforth and which was used earlier in [2]—is to consider a sequence of block-fading channels indexed by their blocklength $J = 1, 2, \cdots$ all of which have the same fading distribution. Then for a given rate of transmission R and specified energy constraints, we derive *firm* upper and lower bounds on the FEP achievable over the channel of finite blocklength J. The limiting behavior of the firm bounds as $J \to \infty$ is then examined. As noted in [2], it is meaningful to study the performance limits as $J \to \infty$ since the block-lengths even in delay-constrained practical systems can be fairly large. For notational convenience, we do not explicitly index the channel matrix by its coherence interval J but simply assume that a code of length J is only used over the channel with coherence interval J [3]. In particular, when we refer to the FEP of "a code of length J" we mean the FEP obtained by using that code over the channel with coherence interval J. Finally, note that for a given channel (with finite blocklength) letting the code length go to infinity takes us into the ergodic regime where the Shannon capacity is positive and which is not of interest here.

---

[3] Recall that all channels in the sequence have the same distribution.

III. PROBLEM FORMULATIONS

A. Outage Probability Definitions

For the model in (1) with only CSIR and an average energy constraint $E[\|\mathbf{X}\|_{\mathrm{F}}^2] \le WJ$, for some constant $W > 0$, the outage probability, denoted by $P_{\mathrm{out}}^{\mathrm{csir}}(R)$, is given by [8]

$$P_{\mathrm{out}}^{\mathrm{csir}}(R) \stackrel{\triangle}{=} \inf_{\mathbf{Q}\succeq\mathbf{0}:\mathrm{tr}(\mathbf{Q})\le W} \Pr\left(\log\left(\left|\mathbf{I}+\mathbf{HQH}^\dagger\right|\right) < R\right), \tag{2}$$

where $\succeq$ denotes the positive semidefinite ordering. Unfortunately, the optimal covariance matrix is not known even in the commonly used single-block i.i.d. Rayleigh fading model. However, for this case Telatar [8] makes a widely accepted conjecture that the optimal strategy is to assign equal power to a subset of the antennas and turn off the rest, i.e., the optimal $\mathbf{Q}$ is of the form $\frac{W}{r}\mathrm{diag}\{\mathbf{1}_r^T, \mathbf{0}_{K-r}^T\},\ 1 \le r \le K$. This conjecture, to the best of our knowledge, has only been proved for systems with one receive antenna ($N = 1$) in [10]. In some works, for instance [11], the covariance matrix is restricted to be a scaled identity matrix and the resulting probability

$$\hat{P}_{\mathrm{out}}^{\mathrm{csir}}(R) \stackrel{\triangle}{=} \Pr\left(\log\left|\mathbf{I}+\frac{W}{K}\mathbf{HH}^\dagger\right| < R\right), \tag{3}$$

is taken to be the outage probability. Clearly, $P_{\mathrm{out}}^{\mathrm{csir}}(R) \le \hat{P}_{\mathrm{out}}^{\mathrm{csir}}(R)$.

On the other hand, for systems with perfect CSIR and perfect CSIT (the combination henceforth referred to as just CSIT), the codebook used can be a function of $\mathbf{H}$. For such systems, a short-term power constraint (STPC), $E[\|\mathbf{X}\|_{\mathrm{F}}^2|\mathbf{H}] \le WJ,\ \forall\ \mathbf{H}$ can be imposed and [2] defines the outage probability to be

$$P_{\mathrm{out}}^{\mathrm{csit-st}}(R) \stackrel{\triangle}{=} \Pr\left(\max_{\mathbf{Q}\succeq\mathbf{0}:\ \mathrm{tr}(\mathbf{Q})\le W} \left\{\log\left|\mathbf{I}+\mathbf{HQH}^\dagger\right|\right\} < R\right). \tag{4}$$

Moreover, a long-term power constraint (LTPC) can also be imposed by letting the transmitter transmit scaled codewords $\sqrt{\gamma(\mathbf{H})}\mathbf{X}$ with $E[\|\mathbf{X}\|_{\mathrm{F}}^2|\mathbf{H}] \le WJ,\ \forall\ \mathbf{H}$ and where $\gamma : \mathbb{C}^{N\times K} \to \mathbb{R}_+$ is any scaling function such that $E[\gamma(\mathbf{H})] \le 1$. Clearly the LTPC is a more relaxed constraint than the STPC. The outage probability for this case is defined in [2] to be

$$P_{\mathrm{out}}^{\mathrm{csit-lt}}(R) \stackrel{\triangle}{=} \min_{\gamma(\mathbf{H})\ge 0:\ E[\gamma(\mathbf{H})]\le 1} \Pr\left(\max_{\mathbf{Q}\succeq\mathbf{0}:\ \mathrm{tr}(\mathbf{Q})\le W} \left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{HQH}^\dagger\right|\right\} < R\right), \tag{5}$$

and an optimal scaling function is provided in [2].

B. A New Class of Codes

In this paper, our interest is in proving both achievability and strong converse theorems for a certain class of codes which we next define. Letting $M_J = \lceil 2^{RJ} \rceil,\ J \geq 1$, denote the number of codewords, we define the following telescoping sets of codes of rate R (in bits per channel use), denoted as $\mathcal{C}_{\mathrm{A}}, \mathcal{C}_{\mathrm{F}}$, and $\mathcal{C}_{\mathrm{P}}$, respectively, as follows:

$$\mathcal{C}_{\mathrm{A}} \triangleq \left\{ \mathcal{X}_J = \{\mathbf{X}^j\}_{j=1}^{M_J},\ \forall J \geq 1 : \mathbf{X}^j \in \mathbb{C}^{K\times J},\ \frac{1}{M_J}\sum_{j=1}^{M_J} \|\mathbf{X}^j\|_{\mathrm{F}}^2 \leq WJ \right\}, \tag{6}$$

$$\mathcal{C}_{\mathrm{F}} \triangleq \left\{ \mathcal{X}_J = \{\mathbf{X}^j\}_{j=1}^{M_J},\ \forall J \geq 1 : \mathbf{X}^j \in \mathbb{C}^{K\times J},\ \|\mathbf{X}^j\|_{\mathrm{F}}^2 \leq WJ,\ \forall\, 1 \leq j \leq M_J \right\}, \tag{7}$$

and

$$\mathcal{C}_{\mathrm{P}} \triangleq \Bigg\{ \mathcal{X}_J = \{\mathbf{A}\mathbf{X}^j\}_{j=1}^{M_J},\ \forall J \geq 1 : \mathbf{A} \in \mathbb{C}^{K\times K},\ \mathrm{tr}(\mathbf{A}\mathbf{A}^\dagger) \leq W,\ \mathbf{X}^j \in \mathbb{C}^{K\times J}, \\ \lambda_{\max}\left(\frac{\mathbf{X}^j(\mathbf{X}^j)^\dagger}{J}\right) \leq 1,\ 1 \leq j \leq M_J \Bigg\}. \tag{8}$$

Note that in (6)–(8) we have used $\mathcal{X}_J$ to denote a space-time codebook of $K \times J$ codeword matrices. The set $\mathcal{C}_{\mathrm{A}}$ is the set of space-time codes of rate R satisfying the average energy constraint whereas the set $\mathcal{C}_{\mathrm{F}}$ is the set of space-time codes of rate R satisfying the maximum energy (Frobenius norm) constraint. The newly introduced class of codes $\mathcal{C}_{\mathrm{P}}$ can also be expressed as

$$\mathcal{C}_{\mathrm{P}} = \left\{ \mathbf{A}\mathcal{C}_{\mathrm{E}} : \mathrm{tr}\left(\mathbf{A}\mathbf{A}^\dagger\right) \leq W \right\},$$

where

$$\mathcal{C}_{\mathrm{E}} \triangleq \left\{ \mathcal{X}_J = \{\mathbf{X}^j\}_{j=1}^{M_J},\ \forall J \geq 1 : \mathbf{X}^j \in \mathbb{C}^{K\times J},\ \lambda_{\max}\left(\frac{\mathbf{X}^j(\mathbf{X}^j)^\dagger}{J}\right) \leq 1,\ 1 \leq j \leq M_J \right\}.$$

The $\mathcal{C}_{\mathrm{E}}$ is the set of codes whose (normalized) spectral norm is bounded by unity and $\mathcal{C}_{\mathrm{P}}$ can be interpreted as its expansion via *precoding* including beamforming and power control. Also, note the telescoping set inclusions

$$\mathcal{C}_{\mathrm{P}} \subset \mathcal{C}_{\mathrm{F}} \subset \mathcal{C}_{\mathrm{A}}.$$

For the special case of a single input, $K = 1$, the sets $\mathcal{C}_{\mathrm{P}}$ and $\mathcal{C}_{\mathrm{F}}$ are identically equal and many arguments in this paper simplify greatly.

C. Achievability of Outage Probability and Best Achievable FEP

Next, we make precise the notions of what it means for the outage probability to be achievable and when there is a strong converse associated with a set of codes. The commonly used definition of achievability for the CSIR-only case is as follows.

Definition 1: A set of codes $\mathcal{C}$ of rate R is said to achieve $P_{\text{out}}^{\text{csir}}(R)$ in (2), if for any $\epsilon > 0$, there is an integer $J_{\text{a}}(\epsilon) \geq 1$, such that for each $J > J_{\text{a}}(\epsilon)$, there exists a code $\mathcal{X}_J \in \mathcal{C}$ yielding a FEP, averaged over the set of channel realizations and denoted by $\Pr\left(\mathcal{E}\,|\mathcal{X}_J\right)$, no greater than $P_{\text{out}}^{\text{csir}}(R) + \epsilon$.

Note that if a set of codes $\mathcal{C}$ of rate R achieves $P_{\text{out}}^{\text{csir}}(R)$, we have that there exists a sequence of codes $\{\mathcal{X}_J\}_{J=1}^{\infty}$ in $\mathcal{C}$ such that $\limsup_{J\to\infty} \Pr(\mathcal{E}|\mathcal{X}_J) \leq P_{\text{out}}^{\text{csir}}(R)$.

In a similar vein, the definition of achievability for the CSIT with LTPC reads as follows:

Definition 2: A set of codes $\mathcal{C}$ of rate R is said to achieve $P_{\text{out}}^{\text{csit}-\text{lt}}(R)$ in (4) if for any $\epsilon > 0$, there is an integer $J_{\text{a}}(\epsilon) \geq 1$ such that for each $J > J_{\text{a}}(\epsilon)$ there exists a scaling function $\gamma(\mathbf{H}) \geq 0$ such that $E[\gamma(\mathbf{H})] \leq 1$ and a coding rule which picks $\mathcal{X}_J(\mathbf{H}) \in \mathcal{C},\ \forall\ \mathbf{H}$, yielding a FEP, averaged over the set of channel realizations and denoted as $E\left[\Pr\left(\mathcal{E}|\mathbf{H}, \mathcal{X}_J(\mathbf{H}), \gamma(\mathbf{H})\right)\right]$ that is no greater than $P_{\text{out}}^{\text{csit}-\text{lt}}(R) + \epsilon$.

Note that in the case of CSIT with STPC, the definition remains the same as above except that now $\gamma(\mathbf{H}) = 1,\ \forall\ \mathbf{H}$.

Given the telescoping set inclusions $\mathcal{C}_{\text{P}} \subset \mathcal{C}_{\text{F}} \subset \mathcal{C}_{\text{A}}$, it is clear that the strongest achievability theorem would be the one proved for $\mathcal{C}_{\text{P}}$ followed by that for $\mathcal{C}_{\text{F}}$ and then $\mathcal{C}_{\text{A}}$.

Let us turn our attention to definitions associated with the *strong converse* results which involve showing that the outage probabilities are a lower bound on the FEPs.

Definition 3: For the CSIR-only case, the strong converse holds for a class of codes $\mathcal{C}$ of rate R in the limit of large frame lengths, if for any $\epsilon > 0$, there is an integer $J_{\text{sc}}(\epsilon) \geq 1$, such that any code $\mathcal{X}_J \in \mathcal{C}$ with frame (codeword) length $J > J_{\text{sc}}(\epsilon)$ has a FEP no less than $P_{\text{out}}^{\text{csir}}(R) - \epsilon$.

Note that if both the achievability of $P_{\text{out}}^{\text{csir}}(R)$ and the corresponding strong converse can be shown for the same set $\mathcal{C}$ of rate R, then $P_{\text{out}}^{\text{csir}}(R)$ is the *best achievable* FEP for $\mathcal{C}$.

Note that if the strong converse holds for a set of codes $\mathcal{C}$ then any sequence of codes $\{\mathcal{X}_J\}_{J=1}^{\infty}$ in $\mathcal{C}$ with increasing (unbounded) framelengths satisfies

$$\liminf_{J\to\infty} \Pr(\mathcal{E}|\mathcal{X}_J) \geq P_{\text{out}}^{\text{csir}}(R). \tag{9}$$

Similarly, the definition of strong converse for the CSIT with LTPC reads as follows:

Definition 4: For the CSIT case with LTPC, the strong converse holds for a set $\mathcal{C}$ of rate R in the limit of large frame lengths, if for any $\epsilon > 0$, there is an integer $J_{\rm sc}(\epsilon) \geq 1$, such that any scaling function $\gamma(\mathbf{H}) \geq 0$ such that $E[\gamma(\mathbf{H})] \leq 1$ and any coding rule which picks $\mathcal{X}_J(\mathbf{H}) \in \mathcal{C},\ \forall\ \mathbf{H},\ J \geq J_{\rm sc}(\epsilon)$, yields an average FEP $E\left[\Pr\left(\mathcal{E}\left|\mathbf{H}, \mathcal{X}_J(\mathbf{H}), \gamma(\mathbf{H})\right.\right)\right]$ that is no less than $P_{\rm out}^{\rm csit-lt}(R) - \epsilon$.

Note again that if achievability of $P_{\rm out}^{\rm csit-lt}(R)$ and the corresponding strong converse can be shown for the same set $\mathcal{C}$ of rate R, then $P_{\rm out}^{\rm csit-lt}(R)$ is the best achievable FEP for $\mathcal{C}$.

Again, in the case of CSIT with STPC, the definition remains the same as above except $\gamma(\mathbf{H}) = 1,\ \forall\ \mathbf{H}$. As discussed in Section I, we will, without loss of generality, assume that the outage probabilities defined in (2), (4) and (5) are strictly positive for the specified rate R and power W.

IV. SUMMARY OF RESULTS

It is not known for MIMO channels whether the strong converse result holds for the set $\mathcal{C}_{\rm A}$, or even $\mathcal{C}_{\rm F}$. The best known result is based on Fano's inequality. In particular, using Fano's inequality, we can lower bound the FEP of a code $\mathcal{X}_J \in \mathcal{C}_{\rm A}$ for the CSIR-only case as

$$\Pr(\mathcal{E}|\mathcal{X}_J) \geq \inf_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ {\rm tr}(\mathbf{Q}) \leq W}} E\left[\left(1 - \frac{1}{RJ} - \frac{\log\left|\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right|}{R}\right)^+\right], \tag{10}$$

where $(x)^+ = \max\{0, x\}$, and for the CSIT case we have

$$\begin{aligned} & E\left[\Pr\left(\mathcal{E}\left|\mathbf{H}, \mathcal{X}_J(\mathbf{H}), \gamma(\mathbf{H})\right.\right)\right] \\ & \quad \geq \inf_{\gamma(\mathbf{H}) \geq 0 : E[\gamma(\mathbf{H})] \leq 1} E\left[\left(1 - \frac{1}{RJ} - \frac{1}{R} \max_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ {\rm tr}(\mathbf{Q}) \leq W}} \log\left|\mathbf{I} + \gamma(\mathbf{H})\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right|\right)^+\right]. \end{aligned} \tag{11}$$

Unfortunately, these lower bounds are strictly less than their respective outage probabilities as $J \to \infty$. In this regard, we note that [12] uses Fano's inequality, but only to conclude that in the CSIR case, the *diversity order* of any code in $\mathcal{C}_{\rm A}$ is no greater than the diversity order of $P_{\rm out}^{\rm csir}(R)$.

Recently, the authors in [4] considered MIMO systems with only CSIR and developed a sphere packing lower bound on the FEP of any code in $\mathcal{C}_{\rm F}$. Unfortunately, the bound developed in [4] is based on some invalid assumptions and approaches $\hat{P}_{\rm out}^{\rm csir}(R)$ in (3) as $J \to \infty$.

Note that since the achievability of $P_{\rm out}^{\rm csir}(R)$ with the set $\mathcal{C}_{\rm F}$ can be readily proved, and since $P_{\rm out}^{\rm csir}(R) \leq \hat{P}_{\rm out}^{\rm csir}(R)$, with strict inequality in many cases, the result in [4] cannot be true. It is shown in Appendix-B that the sphere packing lower bound in [4] is based on an incorrect assumption, and hence is not valid.

The central contribution of this work is that it establishes the outage probability to be the *best achievable* FEP in the large framelength limit for the set $\mathcal{C}_{\rm P}$. It is summarized for the CSIR-only case using Definitions 1 and 3 in the theorem below.

Theorem 1: In the case of CSIR only, the set of codes $\mathcal{C}_{\rm P}$ of rate R defined in (8) achieves the outage probability $P_{\rm out}^{\rm csir}(R)$ in (2) (and hence, so do $\mathcal{C}_{\rm F}$ and $\mathcal{C}_{\rm A}$). Moreover, the corresponding strong converse holds for the same set of codes $\mathcal{C}_{\rm P}$; taken together, these imply that $P_{\rm out}^{\rm csir}(R)$ is the best achievable FEP for the set of codes $\mathcal{C}_{\rm P}$ when only the receiver has CSI.

For the CSIT case with LTPC (STPC), using Definitions 2 and 4, the theorem reads as follows.

Theorem 2: In the case of CSIT with LTPC (STPC), the set of codes $\mathcal{C}_{\rm P}$ of rate R defined in (8) achieves the outage probability $P_{\rm out}^{\rm csit-lt}(R)$ in (5) (resp., $P_{\rm out}^{\rm csit-st}(R)$ in (4), and hence, so do $\mathcal{C}_{\rm F}$ and $\mathcal{C}_{\rm A}$. Moreover, the corresponding strong converse holds for the same set of codes $\mathcal{C}_{\rm P}$; taken together, these imply that $P_{\rm out}^{\rm csit-lt}(R)$ (resp., $P_{\rm out}^{\rm csit-st}(R)$) is the best achievable FEP for the set of codes $\mathcal{C}_{\rm P}$ when both the receiver and the transmitter have CSI under the LTPC constraint.

Note that the theorem for the CSIT with STPC case is included parenthetically and is obtained after setting $\gamma(\mathbf{H}) = 1, \ \forall \ \mathbf{H}$.

In Section V, we use standard random coding arguments to prove the achievability for the set $\mathcal{C}_{\rm P}$. In Section VI, we consider systems with and without CSIT and develop a sphere packing lower bound for the set $\mathcal{C}_{\rm P}$ and using that bound we prove that the strong converse holds for the set $\mathcal{C}_{\rm P}$ in Section VII.

Finally, the notion of universal communications is considered in Section VIII by taking the compound channel view. There, in particular, for positive and finite-valued parameters Δ and a fixed arbitrarily, we define the set

$$\mathcal{A}_{\Delta}^{a} = \left\{ \mathbf{H} \in \mathbb{C}^{N\times K} : \|\mathbf{H}\|_\infty \leq a, \ \log\left|\mathbf{I} + \mathbf{H}\mathbf{H}^\dagger\right| \geq R + \Delta \right\} , \tag{12}$$

and prove the following achievability result for the class of codes $\mathcal{C}_{\rm E}$ whose codewords are bounded in the spectral norm. The role of the parameter a is to make the set $\mathcal{A}_{\Delta}^{a}$ compact. It must be finite but can be taken to be arbitrarily large.

Theorem 3: For any $\epsilon > 0$, $\exists\, \mathcal{X}_J \in \mathcal{C}_{\mathrm{E}}$ such that $\forall\, \mathbf{H} \in \mathcal{A}^{a}_{\Delta},\ \Pr\left(\mathcal{E}\left|\mathbf{H}, \mathcal{X}_J\right.\right) \leq \epsilon$.

The above result strengthens the fundamental result of Root and Varaiya in [9]. The proof extends the elaborate techniques of [9] which prove the corresponding result for the larger class of codes $\mathcal{C}_{\mathrm{F}}$.

Next, we discuss the importance of the above result. Recent works have considered the design of universal codes for the CSIR-only case [13]. Also, design rules along with particular constructions of approximately universal codes have been presented in [14] (see also [15]). These codes are universal in the coarser (high-SNR) diversity-multiplexing tradeoff framework. One aspect that has not been considered in [13] and *cannot* be captured by the framework used in [15], is that of combining a universal code with good precoders. [13] seeks to design codes that are uniformly good over the set $\mathcal{A}^{\infty}_{\Delta}$ and in-fact their stated goal is achieve a FEP close to $\hat{P}^{\mathrm{csir}}_{\mathrm{out}}(R)$ in (3). However, a key observation is that *the problem of designing good codes for arbitrary (but known) fading statistics can be decoupled into the problem of designing universal codes (over $\mathcal{A}^{\infty}_{\Delta}$) and that of designing good precoders*. Put in another way, suppose we choose a precoder $\mathbf{A}:\ \mathrm{tr}(\mathbf{A}\mathbf{A}^{\dagger}) \leq W$, so that the code $\hat{\mathcal{X}}_J = \mathbf{A}\mathcal{X}_J \in \mathcal{C}_{\mathrm{P}}$. Notice that for any $\mathbf{H}$ such that $\log\left|\mathbf{I} + \mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}^{\dagger}\right| \geq R + \Delta$, we have that $\Pr(\mathcal{E}|\hat{\mathcal{X}}_J, \mathbf{H}) \leq \epsilon$. Consequently, the FEP of such a code in the CSIR-only case (where the code $\hat{\mathcal{X}}_J$ is invariant to the channel realization) can be upper-bounded as

$$\Pr(\mathcal{E}|\hat{\mathcal{X}}_J) \leq \epsilon + (1-\epsilon)\Pr\left(\log\left|\mathbf{I} + \mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}^{\dagger}\right| < R + \Delta\right) . \tag{13}$$

Thus by designing a good precoder matched to the given fading distribution (using (2)), we can obtain a performance close to $P^{\mathrm{csir}}_{\mathrm{out}}(R)$ in (2) instead of (3). However, this precoder optimization clearly requires the knowledge of the fading law at the transmitter.

The same idea applies to the scenario with CSIT except that the precoder can depend on the channel realization.

V. PROOFS OF ACHIEVABILITY

In this section we prove the achievability part of Theorems 1 and 2. Let us first consider the CSIR-only case. The key issue here is that since the instantaneous CSI is absent at the transmitter, it cannot adapt or change its code based on the channel realization. Thus, we have to prove that there exists a single code which achieves an error probability close to the outage probability. Then, for given $\epsilon > 0$, we first obtain[4] $\hat{\mathbf{A}}$ such that

$$\Pr\left(\mathcal{O}_{\hat{\mathbf{A}}}\right) \stackrel{\triangle}{=} \Pr\left(\log\left|\mathbf{I}+\mathbf{H}\hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger}\mathbf{H}^{\dagger}\right| < R\right) \le P_{\text{out}}^{\text{csir}}(R) + \epsilon. \tag{14}$$

and $\text{tr}(\hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger}) \le W - \delta$, for some $\delta > 0$.

Let $\mathcal{I}_{\mathcal{O}_{\hat{\mathbf{A}}}}(\mathbf{H})$ denote the indicator function for the event $\left\{\mathbf{H} : \log\left(\left|\mathbf{I}+\mathbf{H}\hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger}\mathbf{H}^{\dagger}\right|\right) < R\right\}$ and let $\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H}) \stackrel{\triangle}{=} 1 - \mathcal{I}_{\mathcal{O}_{\hat{\mathbf{A}}}}(\mathbf{H})$. Using these indicator functions, the average FEP achieved by any code $\mathcal{X}_J \in \mathcal{C}_{\text{P}}$ can be upper bounded as

$$\begin{aligned}\Pr(\mathcal{E}|\mathcal{X}_J) &= E\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\mathcal{I}_{\mathcal{O}_{\hat{\mathbf{A}}}}(\mathbf{H})\right] + E\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right], \\ &\le \Pr\left(\mathcal{O}_{\hat{\mathbf{A}}}\right) + E\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right]. \end{aligned} \tag{15}$$

To bound the term $E\left[\Pr(\mathcal{E}|\mathbf{H},\mathcal{X}_J)\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right]$, we use the standard random coding upper bound with the density

$$f(\mathbf{X}) = \frac{\phi(\mathbf{X})}{\mu}\prod_{j=1}^{J}\frac{1}{\pi^{K}|\boldsymbol{\Sigma}|}\exp\left(-\mathbf{x}_j^{\dagger}\boldsymbol{\Sigma}^{-1}\mathbf{x}_j\right), \tag{16}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_J]$ & $\boldsymbol{\Sigma} = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger}$, and

$$\phi(\mathbf{X}) = \begin{cases} 1 & \frac{\mathbf{X}\mathbf{X}^{\dagger}}{J} \preceq \frac{W}{W-\delta}\hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger} \\ 0 & \text{otherwise} \end{cases}, \tag{17}$$

and $\mu = \Pr\left(\frac{\mathbf{Z}\mathbf{Z}^{\dagger}}{J} \preceq \frac{W}{W-\delta}\hat{\mathbf{A}}\hat{\mathbf{A}}^{\dagger}\right)$, where $\mathbf{Z}$ is a $K \times J$ matrix with i.i.d. $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma})$ columns. Note that the codebook drawn using (16) belongs to $\mathcal{C}_{\text{P}}$. Let $\mathcal{C}_{\text{P}}^{J}$ be the subset of $\mathcal{C}_{\text{P}}$ containing length J codes so that $\mathcal{C}_{\text{P}} = \cup_{J=1}^{\infty}\mathcal{C}_{\text{P}}^{J}$. Then the key step due to Tonelli's theorem [16], is that

$$E_{\mathcal{C}_{\text{P}}^{J}}E_{\mathbf{H}}\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right] = E_{\mathbf{H}}E_{\mathcal{C}_{\text{P}}^{J}}\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right], \tag{18}$$

where $E_{\mathcal{C}_{\text{P}}^{J}}[.]$ denotes the expectation over the set $\mathcal{C}_{\text{P}}^{J}$ using the density in (16). Then, conditioned on $\mathbf{H}$ and letting $\bar{\Pr}\left(\mathcal{E}|\mathbf{H}\right) \stackrel{\triangle}{=} E_{\mathcal{C}_{\text{P}}^{J}}\left[\Pr\left(\mathcal{E}|\mathbf{H},\mathcal{X}_J\right)\right]$, a random coding upper bound can be computed using steps similar to those in [17, 18] as

$$\bar{\Pr}(\mathcal{E}|\mathbf{H}) \le \left(\frac{1}{\mu}\right)^{1+\rho(\mathbf{H})} 2^{-J\rho(\mathbf{H})\left[\log\left|\mathbf{I}+\frac{1}{1+\rho(\mathbf{H})}\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^{\dagger}\right|-R\right]}, \quad \rho(\mathbf{H}) \in [0,1]. \tag{19}$$

---

[4] The existence of such a matrix follows from the continuity $P_{\text{out}}^{\text{csir}}(R)$ in W.

Letting $\hat{\rho}(\mathbf{H}) = \arg\max_{\rho\in[0,1]} \left\{\rho\left[\log\left|\mathbf{I}+\frac{1}{1+\rho}\mathbf{H\Sigma H}^{\dagger}\right| - R\right]\right\}$ and using (19) with (18) and (15), we obtain a random coding upper bound

$$\bar{\mathrm{P}}\mathrm{r}(\mathcal{E}) \leq \Pr(\mathcal{O}_{\hat{\mathbf{A}}}) + E\left[\left(\frac{1}{\mu}\right)^{1+\hat{\rho}(\mathbf{H})} 2^{-J\hat{\rho}(\mathbf{H})\left[\log\left|\mathbf{I}+\frac{1}{1+\hat{\rho}(\mathbf{H})}\mathbf{H\Sigma H}^{\dagger}\right|-R\right]}\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H})\right]. \tag{20}$$

Now, it is known from [17], that for all $\mathbf{H}$ such that $\mathcal{I}_{\mathcal{O}^{c}_{\hat{\mathbf{A}}}}(\mathbf{H}) = 1$, we have

$$\hat{\rho}(\mathbf{H})\left[\log\left|\mathbf{I}+\frac{1}{1+\hat{\rho}(\mathbf{H})}\mathbf{H\Sigma H}^{\dagger}\right| - R\right] \geq 0$$

with equality iff $\log\left|\mathbf{I}+\mathbf{H\Sigma H}^{\dagger}\right| = R$. However, since $\log\left|\mathbf{I}+\mathbf{H\Sigma H}^{\dagger}\right|$ is a continuous random variable, the set $\left\{\mathbf{H} : \log\left|\mathbf{I}+\mathbf{H\Sigma H}^{\dagger}\right| = R\right\}$ is either a set of measure zero or measure one but since the outage probability in (2) is strictly positive, we can conclude that it must be a set of measure zero. Further, using the strong law of large numbers [19], it follows that $\frac{\mathbf{ZZ}^{\dagger}}{J} \to \mathbf{\Sigma}$. Since $\mathbf{\Sigma} \prec \frac{W}{W-\delta}\mathbf{\Sigma}$, we have that $\lim_{J\to\infty}\mu = 1$. Using these facts and the dominated convergence theorem [16], we have the desired result that

$$\limsup_{J\to\infty} \bar{\mathrm{P}}\mathrm{r}(\mathcal{E}) = \Pr(\mathcal{O}_{\hat{\mathbf{A}}}) \leq P_{\mathrm{out}}^{\mathrm{csir}}(R) + \epsilon. \tag{21}$$

This concludes the proof of the achievability part of Theorem 1.

Next, consider the CSIT case with LTPC. Let $\hat{\gamma}(\cdot)$ be an optimal scaling function in the outage minimization in (5) (determined using the results in [2]) when the average transmit power is $W-\delta$ and set $\hat{\mathbf{Q}}(\mathbf{H}) \stackrel{\triangle}{=} \arg\max_{\mathbf{Q}\succeq\mathbf{0},\,\mathrm{tr}(\mathbf{Q})\leq W-\delta}\left\{\log\left|\mathbf{I}+\hat{\gamma}(\mathbf{H})\mathbf{HQH}^{\dagger}\right|\right\}$. Let δ be chosen such that

$$\Pr\left(\log\left|\mathbf{I}+\hat{\gamma}(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right| < R\right) \leq P_{\mathrm{out}}^{\mathrm{csit-lt}}(R) + \epsilon. \tag{22}$$

Now, taking $\mathbf{\Sigma} = \hat{\mathbf{Q}}(\mathbf{H})$ in (16) and $\hat{\gamma}(.)$ to be the scaling function, the random coding upper bound is obtained as

$$\begin{aligned}\bar{\mathrm{P}}\mathrm{r}(\mathcal{E}) &\leq P_{\mathrm{out}}^{\mathrm{csit-lt}}(R) + \epsilon \\ &\quad + E\left[\left(\frac{1}{\mu}\right)^{1+\hat{\rho}(\mathbf{H})} 2^{-J\hat{\rho}(\mathbf{H})\left[\log\left|\mathbf{I}+\frac{1}{1+\hat{\rho}(\mathbf{H})}\hat{\gamma}(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right|-R\right]}\mathbf{1}\left\{\log\left|\mathbf{I}+\hat{\gamma}(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right| \geq R\right\}\right].\end{aligned}$$

At this point, the proof of achievability in Theorem 2 follows after noting that $\lim_{J\to\infty}\mu = 1$ (for each $\mathbf{H}$) and using the dominated convergence theorem. The proof of achievability for the CSIT case with STPC follows after taking $\hat{\gamma}(\mathbf{H}) = 1,\ \forall\ \mathbf{H}$ in the above argument.

VI. SPHERE PACKING LOWER BOUNDS

In this section we derive our sphere packing lower bounds on the achievable error probabilities. Note that these are firm bounds which are applicable for finite code lengths. The asymptotic analysis of these bounds (as $J \to \infty$) is undertaken in the next section and it yields the desired strong converses of Theorems 1 and 2.

Lemma 1: For the CSIR-only case, the error probability of any code $\mathcal{X}_J \in \mathcal{C}_{\mathrm{P}}$, is lower bounded by $\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csir}}(\mathrm{W},\mathrm{R},\mathrm{J})$, where

$$\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csir}}(\mathrm{W},\mathrm{R},\mathrm{J}) \triangleq \inf_{\mathbf{A}\in\mathbb{C}^{\mathrm{K}\times\mathrm{K}}:\mathrm{tr}(\mathbf{A}\mathbf{A}^\dagger)\leq\mathrm{W}} \mathrm{E}\left[\left(\sup_{\delta\in\,\mathbb{R}_+}\left\{\exp\left(-\left(\mathrm{r}_{\mathrm{J}}^{\delta}(\mathbf{HA})\right)^2\right)\sum_{\mathrm{j}=0}^{\mathrm{NJ}-1}\frac{(\mathrm{r}_{\mathrm{J}}^{\delta}(\mathbf{HA}))^{2\mathrm{j}}}{\mathrm{j}!} - \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\right\}\right)^+\right], \quad (23)$$

where

$$\left(r_J^{\delta}(\mathbf{HA})\right)^2 \triangleq NJ(1+\delta)2^{-R/N}\left|\mathbf{I}+\frac{1}{1+\delta}\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right|^{1/N} \quad (24)$$

and

$$\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA}) \triangleq \Pr\left(\sum_{\mathrm{k}=1}^{\mathrm{m}}\frac{\|\mathbf{v}_{\mathrm{k}}+\mathrm{s}_{\mathrm{k}}\mathbf{1}_{\mathrm{J}}\|^2}{\mathrm{NJ}(1+\delta+\mathrm{s}_{\mathrm{k}}^2)} + \sum_{\mathrm{k}=\mathrm{m}+1}^{\mathrm{N}}\frac{\|\mathbf{v}_{\mathrm{k}}\|^2}{\mathrm{NJ}(1+\delta)} > 1|\mathbf{HA}\right), \quad (25)$$

with $\mathbf{v}_k,\ 1\leq k\leq N$ denoting the k^{th} row of the noise matrix $\mathbf{V}$ and $\{s_k\}_{k=1}^{m\triangleq\min(N,K)}$ the m largest singular values of $\mathbf{HA}$.

Next, for the CSIT-STPC case, the error probability achieved by any coding rule which picks $\mathcal{X}_J(\mathbf{H})\in\mathcal{C}_{\mathrm{P}},\ \forall\ \mathbf{H}$, is lower bounded by

$$\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-st}}(\mathrm{W},\mathrm{R},\mathrm{J}) \triangleq \mathrm{E}\left[\inf_{\mathbf{A}\in\mathbb{C}^{\mathrm{K}\times\mathrm{K}}:\mathrm{tr}(\mathbf{A}\mathbf{A}^\dagger)\leq\mathrm{W}}\left(\sup_{\delta\,\in\mathbb{R}_+}\left\{\exp\left(-(\mathrm{r}_{\mathrm{J}}^{\delta}(\mathbf{HA}))^2\right)\sum_{\mathrm{j}=0}^{\mathrm{NJ}-1}\frac{\left(\mathrm{r}_{\mathrm{J}}^{\delta}(\mathbf{HA})\right)^{2\mathrm{j}}}{\mathrm{j}!} - \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\right\}\right)^+\right]. \quad (26)$$

Finally, with CSIT-LTPC the error probability achieved by any scaling function $\gamma(\mathbf{H})\geq 0:$ $E[\gamma(\mathbf{H})]\leq 1$ and any coding rule which picks $\mathcal{X}_J(\mathbf{H})\in\mathcal{C}_{\mathrm{P}},\ \forall\ \mathbf{H}$ is is lower bounded by $\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-lt}}(\mathrm{W},\mathrm{R},\mathrm{J})$, which equals

$$\inf_{\gamma(\mathbf{H})\geq 0:\ E[\gamma(\mathbf{H})]\leq 1} E\left[\inf_{\mathbf{A}\in\mathbb{C}^{K\times K}:\mathrm{tr}(\mathbf{A}\mathbf{A}^\dagger)\leq W}\left(\sup_{\delta\in\,\mathbb{R}_+}\left\{\exp\left(-(r_J^{\delta}(\sqrt{\gamma(\mathbf{H})}\mathbf{HA}))^2\right)\right.\right.\right.$$

$$\times \sum_{j=0}^{NJ-1} \frac{\left(r_J^\delta(\sqrt{\gamma(\mathbf{H})}\mathbf{HA})\right)^{2j}}{j!} - \mathrm{Pe}^{\mathrm{ub}}\left(\sqrt{\gamma(\mathbf{H})}\mathbf{HA}\right)\Big\}\Big)^{+}\Big]. \tag{27}$$

Proof: We consider any code $\mathcal{X}_J = \{\mathbf{AX}^j\}_{j=1}^{M_J} \in \mathcal{C}_{\mathrm{P}}$ so that the received matrix can be written as

$$\mathbf{Y} = \mathbf{HAX} + \mathbf{V}\ . \tag{28}$$

For the given $\mathbf{HA}$ and some $\delta \geq 0$, we define an bounding region

$$S_J^\delta(\mathbf{HA}) \overset{\triangle}{=} \left\{\mathbf{Z} \in \mathbb{C}^{N\times J} : \mathrm{tr}\left(\mathbf{Z}^\dagger\left(NJ(1+\delta)\mathbf{I} + NJ\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right)^{-1}\mathbf{Z}\right) \leq 1\right\}. \tag{29}$$

Now letting $\mathcal{E}^c$ denote the complement of the frame (codeword) error event $\mathcal{E}$, we can lower bound the conditional error probability as

$$\Pr\left(\mathcal{E}\,|\mathbf{HA},\mathcal{X}_J\right) \geq 1 - \Pr\left(\mathcal{E}^c, \mathbf{Y} \in S_J^\delta(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right) - \Pr\left(\mathbf{Y} \notin S_J^\delta(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right). \tag{30}$$

We note that the bounded region $S_J^\delta(\mathbf{HA})$ is a hyper-ellipsoid of volume

$$\mathrm{Vol}\left(S_J^\delta(\mathbf{HA})\right) = \frac{\pi^{NJ}}{(NJ)!}(NJ(1+\delta))^{NJ}\left|\mathbf{I} + \frac{1}{1+\delta}\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right|^J. \tag{31}$$

Next, we have

$$\begin{aligned}\Pr\left(\mathcal{E}^c, \mathbf{Y} \in S_J^\delta(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right) &= \frac{1}{M_J}\sum_{j=1}^{M_J}\Pr\left(\mathcal{E}^c, \mathbf{Y} \in S_J^\delta(\mathbf{HA})\,\middle|\mathbf{HA},\mathbf{X}^j\right) \\ &= \frac{1}{M_J}\sum_{j=1}^{M_J}\int_{\mathcal{V}_j(\mathbf{HA})\cap S_J^\delta(\mathbf{HA})} f\left(\mathbf{Y}|\mathbf{HA},\mathbf{X}^j\right)\, d\mathbf{Y}\,, \end{aligned} \tag{32}$$

where, conditioned on $\mathbf{HA}$, $\mathcal{V}_j(\mathbf{HA})$ denotes the Voronoi (decision) region of $\mathbf{X}^j$. Using the fact that $f\left(\mathbf{Y}\,|\mathbf{HA},\mathbf{X}^j\right)$ is a (white) Gaussian density with mean $\mathbf{HAX}^j$, [4] shows that

$$\begin{aligned}&\int_{\mathcal{V}_j(\mathbf{HA})\cap S_J^\delta(\mathbf{HA})} f\left(\mathbf{Y}\,\middle|\mathbf{HA},\mathbf{X}^j\right)\, d\mathbf{Y} \\ &\qquad \leq 1 - \exp\left(-\left(r_J^\delta(\mathbf{HA},\mathbf{X}^j)\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^\delta(\mathbf{HA},\mathbf{X}^j)\right)^{2j}}{j!}\end{aligned} \tag{33}$$

where $r_J^\delta\left(\mathbf{HA},\mathbf{X}^j\right)$ is the radius of a NJ-hypersphere having the same volume as $\mathcal{V}_j(\mathbf{HA}) \cap S_J^\delta(\mathbf{HA})$. Moreover, since $\sum_{j=1}^{M_J}\mathrm{Vol}(\mathcal{V}_j(\mathbf{HA}) \cap S_J^\delta(\mathbf{HA})) = \mathrm{Vol}(S_J^\delta(\mathbf{HA}))$, it is proved in [4] that

$$\frac{1}{M_J}\sum_{j=1}^{M_J}\left(1-\exp\left(-\left(r_J^{\delta}(\mathbf{HA},\mathbf{X}^j)\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(\left(r_J^{\delta}(\mathbf{HA},\mathbf{X}^j)\right)^2\right)^j}{j!}\right)$$
$$\leq 1-\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^{\delta}(\mathbf{HA})\right)^{2j}}{j!}\,, \tag{34}$$

where $r_J^{\delta}(\mathbf{HA})$ is given by (24). Note that $r_J^{\delta}(\mathbf{HA})$ is the radius of a $NJ-$hypersphere whose volume equals $\text{Vol}(S_J^{\delta}(\mathbf{HA}))/M_J$. From (34), we have that the best Voronoi regions which are contained in the bounding set (29) of volume (31), are hyper-spheres of equal volume. Hence, using (34) and (33) in (32), the conditional error probability is lower bounded by a difference of two terms as follows

$$\Pr\left(\mathcal{E}^c,\mathbf{Y}\in S_J^{\delta}(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right)\leq 1-\exp\left(-(r_J^{\delta}(\mathbf{HA}))^2\right)\sum_{j=0}^{NJ-1}\frac{(r_J^{\delta}(\mathbf{HA}))^{2j}}{j!}. \tag{35}$$

Using (35) in (30), we have that

$$\Pr\left(\mathcal{E}\,|\mathbf{HA},\mathcal{X}_J\right)\geq\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^{\delta}(\mathbf{HA})\right)^{2j}}{j!}-\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right). \tag{36}$$

Now to bound the term $\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right)$, we first expand it as

$$\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right)=\frac{1}{M_J}\sum_{j=1}^{M_J}\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\,\middle|\mathbf{HA},\mathbf{X}^j\right). \tag{37}$$

Then we consider the term $\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\,|\mathbf{HA},\mathbf{X}^j\right)$ and since the following analysis is applicable to all $\mathbf{X}^j,\ 1\leq j\leq M_J$, we drop the superscript j. We next write the singular value decompositions of $\mathbf{HA}$ and $\mathbf{X}$ as

$$\mathbf{HA}=\mathbf{RSW}^{\dagger},\quad \mathbf{X}=\mathbf{F}\mathbf{\Lambda}\mathbf{G}^{\dagger}\,, \tag{38}$$

and define the $K\times K$ unitary matrix $\mathbf{Q}=\mathbf{W}^{\dagger}\mathbf{F}$ and let $\mathbf{q}_k,\ 1\leq k\leq K$ denote the k^{th} row of $\mathbf{Q}$. Then, we have

$$\Pr\left(\mathbf{Y}\notin S_J^{\delta}(\mathbf{HA})\middle|\mathbf{HA},\mathbf{X}\right)=\Pr\left(\text{tr}\left(\mathbf{Y}^{\dagger}\left(NJ(1+\delta)\mathbf{I}+NJ\mathbf{HAA}^{\dagger}\mathbf{H}^{\dagger}\right)^{-1}\mathbf{Y}\right)>1\middle|\mathbf{HA},\mathbf{X}\right)$$
$$=\Pr\left(\text{tr}\left(\left(NJ(1+\delta)\mathbf{I}+NJ\mathbf{HAA}^{\dagger}\mathbf{H}^{\dagger}\right)^{-1}(\mathbf{HAX}+\mathbf{V})\,(\mathbf{HAX}+\mathbf{V})^{\dagger}\right)>1\middle|\mathbf{HA},\mathbf{X}\right)$$

$$\begin{aligned}
&\overset{(a)}{=} \Pr\left(\mathrm{tr}\left(\left(NJ(1+\delta)\mathbf{I}+NJ\mathbf{S}\mathbf{S}^\dagger\right)^{-1}\left(\mathbf{S}\mathbf{Q}\mathbf{\Lambda}\mathbf{G}^\dagger+\mathbf{V}\right)\left(\mathbf{S}\mathbf{Q}\mathbf{\Lambda}\mathbf{G}^\dagger+\mathbf{V}\right)^\dagger\right)>1\middle|\mathbf{HA},\mathbf{X}\right)\\
&\overset{(b)}{=} \Pr\left(\mathrm{tr}\left(\left(NJ(1+\delta)\mathbf{I}+NJ\mathbf{S}\mathbf{S}^\dagger\right)^{-1}\left(\mathbf{S}\mathbf{Q}\mathbf{\Lambda}+\mathbf{V}\right)\left(\mathbf{S}\mathbf{Q}\mathbf{\Lambda}+\mathbf{V}\right)^\dagger\right)>1\middle|\mathbf{HA},\mathbf{X}\right)\\
&= \Pr\left(\sum_{k=1}^{m}\frac{\|\mathbf{v}_k+s_k\mathbf{q}_k\mathbf{\Lambda}\|^2}{NJ(1+\delta+s_k^2)}+\sum_{k=m+1}^{N}\frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)}>1\middle|\mathbf{HA},\mathbf{X}\right)\ , \qquad (39)
\end{aligned}$$

where (a) and (b) follow since the additive noise is independent of $\mathbf{X}$ and $\mathbf{HA}$, and is unitarily invariant. Note that since $\lambda_{\max}(\mathbf{X}\mathbf{X}^\dagger)\le J$ and $\|\mathbf{q}_k\|^2=1$, we have $\|s_k\mathbf{q}_k^\dagger\mathbf{\Lambda}\|^2\le s_k^2J,\ 1\le k\le m$. Using this with a *stochastic ordering* result given in Appendix -A, we obtain

$$\begin{aligned}
&\Pr\left(\sum_{k=1}^{m}\frac{\|\mathbf{v}_k+s_k\mathbf{q}_k\mathbf{\Lambda}\|^2}{NJ(1+\delta+s_k^2)}+\sum_{k=m+1}^{N}\frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)}>1\middle|\mathbf{X},\mathbf{HA}\right)\le\\
&\qquad\Pr\left(\sum_{k=1}^{m}\frac{\|\mathbf{v}_k+s_k\mathbf{1}_J\|^2}{NJ(1+\delta+s_k^2)}+\sum_{k=m+1}^{N}\frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)}>1\middle|\mathbf{HA}\right). \qquad (40)
\end{aligned}$$

Then using the upper bound (40) and the result in (39) in (37), we obtain the upper-bound

$$\Pr\left(\mathbf{Y}\notin S_J^\delta(\mathbf{HA})\,|\mathbf{HA},\mathcal{X}_J\right)\le \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA}), \qquad (41)$$

where $\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})$ given in (25).

The sphere packing lower bounds can be obtained by first using the upper bound (41) in (36) to get

$$\Pr(\mathcal{E}|\mathbf{HA},\mathcal{X}_J)\ge\left(\sup_{\delta\in\mathbb{R}_+}\left\{\exp\left(-(r_J^\delta(\mathbf{HA}))^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^\delta(\mathbf{HA})\right)^{2j}}{j!}-\mathrm{Pe}^{\mathrm{ub}}\left(\mathbf{HA}\right)\right\}\right)^{+}. \qquad (42)$$

Now, the sphere-packing lower bound for the CSIR-only case given in (23) can be obtained by minimizing the expected value of the lower bound in (42) over $\mathbf{A}$. For the perfect CSIT case with STPC, the lower bound given in (26) is obtained by using the fact that due to the availability of CSIT, the precoder $\mathbf{A}$ can be channel dependent. Finally, for the perfect CSIT case with LTPC, the bound given in (27) follows after minimizing (42) over $\mathbf{A}$ for each $\mathbf{H}$ and then further minimizing over all possible scaling functions. ■

To evaluate the term $\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})$ given in (25), note that conditioned on $\mathbf{HA}$, the quadratic form

$$\sum_{k=1}^{m}\frac{\|\mathbf{v}_k+s_k\mathbf{1}_J\|^2}{NJ(1+\delta+s_k^2)}+\sum_{k=m+1}^{N}\frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)}$$

is a non-central definite quadratic form in complex normal random variables. Several expressions on the cdf of such a quadratic form along with efficient numerical techniques to evaluate them are given in [20]. Here, we content ourselves with the Chebyshev and Chernoff upper bounds. For any $\beta \in \mathbb{R}$ we have that [21]

$$\begin{aligned} E\left[\|\mathbf{v}_k + \beta \mathbf{1}_J\|^2\right] &= J + J\beta^2, \\ \text{Var}\left(\|\mathbf{v}_k + \beta \mathbf{1}_J\|^2\right) &= J + 2J\beta^2, \quad \text{and} \\ E[\exp(s\|\mathbf{v}_k + \beta \mathbf{1}_J\|^2)] &= \frac{1}{(1-s)^J} \exp\left(\frac{sJ\beta^2}{1-s}\right), \ \ 0 \le s < 1. \end{aligned}$$

Using (43), for any $\delta > 0$, we first obtain a Chebyshev upper bound

$$\begin{aligned} &\Pr\left(\sum_{k=1}^{m} \frac{\|\mathbf{v}_k + s_k \mathbf{1}_J\|^2}{NJ(1+\delta+s_k^2)} + \sum_{k=m+1}^{N} \frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)} > 1 \middle| \mathbf{HA}\right) \\ &\quad \le \frac{\sum_{k=1}^{m} \frac{\text{Var}\left(\|\mathbf{v}_k + s_k \mathbf{1}_J\|^2\right)}{(NJ(1+\delta+s_k^2))^2} + \sum_{k=m+1}^{N} \frac{\text{Var}\left(\|\mathbf{v}_k\|^2\right)}{(NJ(1+\delta))^2}}{\left(1 - \sum_{k=1}^{m} \frac{E[\|\mathbf{v}_k + s_k \mathbf{1}_J\|^2]}{NJ(1+\delta+s_k^2)} - \sum_{k=m+1}^{N} \frac{E[\|\mathbf{v}_k\|^2]}{NJ(1+\delta)}\right)^2} \\ &\quad = \frac{1}{J\left(N - \sum_{k=1}^{m} \frac{1+s_k^2}{1+\delta+s_k^2} - \frac{(N-m)}{(1+\delta)}\right)^2} \left(\sum_{k=1}^{m} \frac{1+2s_k^2}{(1+\delta+s_k^2)^2} + \frac{N-m}{(1+\delta)^2}\right). \end{aligned} \tag{43}$$

Letting $e_k = N(1+\delta+s_k^2), \ 1 \le k \le m$, we get the following Chernoff upper bound

$$\begin{aligned} &\Pr\left(\sum_{k=1}^{m} \frac{\|\mathbf{v}_k + s_k \mathbf{1}_J\|^2}{NJ(1+\delta+s_k^2)} + \sum_{k=m+1}^{N} \frac{\|\mathbf{v}_k\|^2}{NJ(1+\delta)} > 1 \,|\mathbf{HA}\right) \\ &\quad \le \min_{s \in [0, aJ)} \left\{ \frac{1}{\exp(s)} \prod_{k=1}^{m} E\left[\exp\left(s(NJ(1+\delta+s_k^2))^{-1} \|\mathbf{v}_k + s_k \mathbf{1}_J\|^2\right)\right] \right. \\ &\qquad\qquad \left. \times E\bigg[\exp\left(\frac{s\sum_{k=m+1}^{N} \|\mathbf{v}_k\|^2}{NJ(1+\delta)}\right)\bigg] \right\} \\ &\quad \overset{(b)}{=} \left(\min_{x \in [0,a)} \left\{ \frac{1}{\exp(x)} \frac{1}{\left(1 - \frac{x}{N(1+\delta)}\right)^{(N-m)}} \prod_{k=1}^{m} \frac{1}{\left(1 - \frac{x}{e_k}\right)} \exp\left(\frac{x s_k^2}{e_k(1 - \frac{x}{e_k})}\right) \right\}\right)^J, \end{aligned} \tag{44}$$

where (b) follows by setting $x = s/J$, and where $a = \min\{e_k\}$ if $m = N$ and $a = N(1+\delta)$ for $m < N$.

VII. PROOF OF STRONG CONVERSE

Note that the strong converse will be proved in each of the three cases if we can show the following inequalities

$$\liminf_{J\to\infty} \mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csir}}(\mathrm{W},\mathrm{R},\mathrm{J}) \geq \inf_{\mathbf{Q}\succeq\mathbf{0}:\mathrm{tr}(\mathbf{Q})\leq W} \Pr\left(\log\left(\left|\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\right|\right) < R\right), \tag{45}$$

$$\liminf_{J\to\infty} \mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-st}}(\mathrm{W},\mathrm{R},\mathrm{J}) \geq \Pr\left(\max_{\mathbf{Q}\succeq\mathbf{0}:\mathrm{tr}(\mathbf{Q})\leq W}\left\{\log\left|\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\right|\right\} < R\right), \tag{46}$$

and

$$\liminf_{J\to\infty} \mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-lt}}(\mathrm{W},\mathrm{R},\mathrm{J}) \geq \inf_{\gamma(\mathbf{H})\geq 0:\ E[\gamma(\mathbf{H})]\leq 1} \Pr\left(\max_{\mathbf{Q}\succeq\mathbf{0}:\mathrm{tr}(\mathbf{Q})\leq W}\left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\right|\right\} < R\right), \tag{47}$$

respectively. We will consider the CSIR and CSIT cases in separate subsections.

A. *CSIR-only*

We start with the CSIR-only case. We fix $\delta = 2^{\Delta/2N}-1$ for some $\Delta > 0$ and recall that for any real-valued scalar $(x)^{+} = \max\{0, x\}$. Then note that for any positive δ

$$\begin{aligned}
&\left(\sup_{\delta\in\mathbb{R}_+}\left\{\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{(r_J^{\delta}(\mathbf{HA}))^{2j}}{j!} - \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\right\}\right)^{+}\\
&\geq \left(\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{(r_J^{\delta}(\mathbf{HA}))^{2j}}{j!} - \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\right)^{+}\\
&\geq \left(\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{(r_J^{\delta}(\mathbf{HA}))^{2j}}{j!} - \mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\right)\times\\
&\qquad \mathbf{1}\Big\{\log\left|\mathbf{I}+\mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}^{\dagger}\right| < R-\Delta\Big\}
\end{aligned}$$

so that we can lower bound $\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csir}}(\mathrm{W},\mathrm{R},\mathrm{J})$ in (23) by

$$\begin{aligned}
&\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{A}\mathbf{A}^{\dagger})\leq W} E\left[\exp\left(-\left(r_J^{\delta}(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^{\delta}(\mathbf{HA})\right)^{2j}}{j!}\mathbf{1}\Big\{\log\left|\mathbf{I}+\mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}^{\dagger}\right| < R-\Delta\Big\}\right]\\
&\qquad - \sup_{\mathbf{A}:\mathrm{tr}\left(\mathbf{A}\mathbf{A}^{\dagger}\right)\leq W} E\left[\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\mathbf{1}\Big\{\log\left|\mathbf{I}+\mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}^{\dagger}\right| < \mathrm{R}-\Delta\Big\}\right].
\end{aligned} \tag{48}$$

We next consider the first term in (48). Note that with i.i.d. $Z_k \sim \mathcal{CN}(0,1),\ 1 \le k \le NJ$, we have

$$\Pr\left(\sum_{k=1}^{NJ} |Z_k|^2 \ge \left(r_J^\delta(\mathbf{HA})\right)^2\right) \;=\; \exp\left(-\left(r_J^\delta(\mathbf{HA})\right)^2\right) \sum_{j=0}^{NJ-1} \frac{\left(r_J^\delta(\mathbf{HA})\right)^{2j}}{j!}. \tag{49}$$

Furthermore, for any $\mathbf{H}$ such that $\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta$, we have that

$$\left|\mathbf{I}+\frac{1}{1+\delta}\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right|^{1/N} \le \left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right|^{1/N} \le 2^{(R-\Delta)/N} \tag{50}$$

and since $1+\delta = 2^{\Delta/2N}$, from (24) we have that

$$\left(r_J^\delta(\mathbf{HA})\right)^2 \le NJ2^{-\Delta/2N} < NJ, \tag{51}$$

so that

$$\begin{aligned}\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{AA}^\dagger)\le W} E&\left[\exp\left(-\left(r_J^\delta(\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^\delta(\mathbf{HA})\right)^{2j}}{j!}\mathbf{1}\Big\{\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta\Big\}\right]\\ &\ge \Pr\left(\sum_{k=1}^{NJ}|Z_k|^2 \ge NJ2^{-\Delta/2N}\right)\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{AA}^\dagger)\le W}\Pr\Big(\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta\Big).\end{aligned} \tag{52}$$

Next, using weak law of large numbers, we have

$$\begin{aligned}\liminf_{J\to\infty}\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{AA}^\dagger)\le W} E\Bigg[\exp\left(-\left(r_J^\delta(\mathbf{HA})\right)^2\right)&\sum_{j=0}^{NJ-1}\frac{\left(r_J^\delta(\mathbf{HA})\right)^{2j}}{j!}\\ &\times\mathbf{1}\Big\{\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta\Big\}\Bigg]\\ \ge \inf_{\mathbf{A}:\mathrm{tr}\left(\mathbf{AA}^\dagger\right)\le W}&\Pr\Big(\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta\Big).\end{aligned} \tag{53}$$

Now, for the second term in (48), we upper bound the Chebyshev bound in (43) as

$$\frac{1}{J\left(N-\sum_{k=1}^m\frac{1+s_k^2}{1+\delta+s_k^2}-\frac{N-m}{1+\delta}\right)^2}\left(\sum_{k=1}^m\frac{1+2s_k^2}{(1+\delta+s_k^2)^2}+\frac{N-m}{(1+\delta)^2}\right)\overset{\text{(a)}}{\le}\frac{b}{J}\,, \tag{54}$$

where $b>0$ is some finite constant with respect to J and (a) follows since Δ,δ are both strictly positive and $\{s_k^2\}$ are bounded when $\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < R-\Delta$. From (54), we obtain

$$\sup_{\mathbf{A}:\mathrm{tr}(\mathbf{AA}^\dagger)\le W} E\left[\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\mathbf{1}\left\{\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < \mathrm{R}-\Delta\right\}\right] \le \frac{b}{J}\,, \tag{55}$$

so that

$$\lim_{J\to\infty} \sup_{\mathbf{A}:\mathrm{tr}(\mathbf{A}\mathbf{A}^\dagger)\le W} E\left[\mathrm{Pe}^{\mathrm{ub}}(\mathbf{HA})\mathbf{1}\left\{\log\left|\mathbf{I}+\mathbf{HAA}^\dagger\mathbf{H}^\dagger\right| < \mathrm{R}-\Delta\right\}\right] = 0. \tag{56}$$

Finally, using (53) and (56) with the lower bound in (48) and recalling that the outage probability is continuous in R, we can conclude that (45) holds, and we are done.

As an aside, note that (56) can be inferred from the Chernoff upper bounds which in fact shows that this term *decays exponentially in* J. To see this, first note that for a non-negative continuous random variable X and any $s \in \mathbb{R}$, we have

$$\begin{aligned}\frac{d^2}{ds^2}\left(\ln(E[\exp(sX)])\right) &= \frac{E[\exp(sX)]E[X^2\exp(sX)] - (E[X\exp(sX)])^2}{(E[\exp(sX)])^2} \\ &\overset{\text{(c)}}{\geq} 0\,, \end{aligned} \tag{57}$$

where (c) follows after setting $Y = \exp(sX/2)$ and $Z = X\exp(sX/2)$ and invoking the Cauchy-Schwartz inequality that $(E[YZ])^2 \leq E[Y^2]E[Z^2]$. Then, since $\ln(E[\exp(sX)])$ is convex in s, for any $\mathbf{A} \in \mathbb{C}^{K\times K} : \mathrm{tr}(\mathbf{AA}^\dagger) \leq W$ the minimization in (44) is equivalent to the convex optimization problem

$$\min_{x\in[0,a)}\left\{\ln\left(\frac{1}{\exp(x)}\frac{1}{\left(1-\frac{x}{N(1+\delta)}\right)^{(N-m)}}\prod_{k=1}^{m}\frac{1}{\left(1-\frac{x}{e_k}\right)}\exp\left(\frac{xs_k^2}{e_k(1-\frac{x}{e_k})}\right)\right)\right\}. \tag{58}$$

Hence the KKT conditions [22] are both necessary and sufficient. Let $\ln(g(x))$ denote the objective function in (58). Then using the KKT conditions it is readily seen that when $\delta > 0$ and $\{s_k^2\}$ are bounded, $x = 0$ is *not* a KKT point and hence at any optimal point $\hat{x}$ we have $\ln(g(\hat{x})) < \ln(g(0)) = 0$ which implies that $g(\hat{x}) < 1$ and hence $(g(\hat{x}))^J \to 0$ as $J \to \infty$. Thus, the second term in (48) decays exponentially to zero as $J \to \infty$.

B. CSIT

For brevity, we prove (47) in the following and omit the proof of (46) which follows after minor changes. We again fix Δ and take δ as in the previous section and lower bound $\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-lt}}(\mathrm{W},\mathrm{R},\mathrm{J})$ in (27) by

$$\inf_{\gamma(\mathbf{H})\geq 0:\, E[\gamma(\mathbf{H})]\leq 1}\left\{E\left[\left(\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{AA}^\dagger)\leq W}\exp\left(-\left(r_J^\delta(\sqrt{\gamma(\mathbf{H})}\mathbf{HA})\right)^2\right)\sum_{j=0}^{NJ-1}\frac{\left(r_J^\delta\left(\sqrt{\gamma(\mathbf{H})}\mathbf{HA}\right)\right)^{2j}}{j!}\right)\right.\right.$$

$$\times\mathbf{1}\left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right|<R-\Delta\right\}\Bigg]$$

$$-E\left[\left(\sup_{\mathbf{A}:\mathrm{tr}(\mathbf{A}\mathbf{A}^{\dagger})\le W}\mathrm{Pe}^{\mathrm{ub}}\left(\sqrt{\gamma(\mathbf{H})}\mathbf{H}\mathbf{A}\right)\right)\mathbf{1}\left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right|<\mathrm{R}-\Delta\right\}\right]\Bigg\}, \qquad (59)$$

where

$$\hat{\mathbf{Q}}(\mathbf{H})=\arg\max_{\mathbf{Q}\succeq\mathbf{0},\ \mathrm{tr}(\mathbf{Q})\le W}\left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\mathbf{Q}\mathbf{H}\right|\right\}.$$

Now note that for any $\mathbf{H}$ such that $\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}\right|<R-\Delta$, we have that

$$\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{H}\right|<R-\Delta,\ \ \forall\ \mathbf{A}:\ \mathrm{tr}\left(\mathbf{A}\mathbf{A}^{\dagger}\right)\le W.$$

Employing the same arguments used to prove (52) and (54), albeit after replacing $\mathbf{H}$ with $\sqrt{\gamma(\mathbf{H})}\mathbf{H}$, we obtain the inequalities

$$E\left[\left(\inf_{\mathbf{A}:\mathrm{tr}(\mathbf{A}\mathbf{A}^{\dagger})\le W}\exp\left(-(r_J^{\delta}(\sqrt{\gamma(\mathbf{H})}\mathbf{H}\mathbf{A}))^2\right)\sum_{j=0}^{NJ-1}\frac{(r_J^{\delta}(\sqrt{\gamma(\mathbf{H})}\mathbf{H}\mathbf{A}))^{2j}}{j!}\right)\right.$$
$$\left.\times\mathbf{1}\left\{\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right|<R-\Delta\right\}\right]$$
$$\ge\Pr\left(\sum_{k=1}^{NJ}|Z_k|^2\ge NJ2^{-\Delta/2N}\right)\Pr\left(\log\left|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}\right|<R-\Delta\right) \qquad (60)$$

and

$$E\left[\left(\sup_{\mathbf{A}:\mathrm{tr}(\mathbf{A}\mathbf{A}^{\dagger})\le W}\mathrm{Pe}^{\mathrm{ub}}\left(\sqrt{\gamma(\mathbf{H})}\mathbf{H}\mathbf{A}\right)\right)\mathbf{1}\left\{\log|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}|<R-\Delta\right\}\right]$$
$$\le\frac{b}{J}\Pr\left(\log|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}|<R-\Delta\right), \qquad (61)$$

so that

$$\mathrm{Pe}_{\mathrm{sp}}^{\mathrm{csit-lt}}(\mathrm{W},\mathrm{R},\mathrm{J})\ge\left(\Pr\left(\sum_{\mathrm{k}=1}^{\mathrm{NJ}}|\mathrm{Z_k}|^2\ge\mathrm{NJ}2^{-\Delta/2\mathrm{N}}\right)-\frac{\mathrm{b}}{\mathrm{J}}\right)$$
$$\times\inf_{\gamma(\mathbf{H})\ge0:\ E[\gamma(\mathbf{H})]\le1}\left\{\Pr\left(\log|\mathbf{I}+\gamma(\mathbf{H})\mathbf{H}\hat{\mathbf{Q}}(\mathbf{H})\mathbf{H}^{\dagger}|<R-\Delta\right)\right\}. \qquad (62)$$

Taking the limit $J\to\infty$ in (62) and again using the continuity of the corresponding outage probability in R, we conclude that (47) holds.

VIII. PRECODING THE UNIVERSAL CODE

In this section we prove that the set $\mathcal{C}_{\mathrm{E}}$ in (8) contains universal codes that can communicate reliably over any channel realization for which the mutual information exceeds the information rate of the code.

To prove Theorem 3, we follow the approach of Root and Varaiya in [9] where it is shown that there exists a code $\mathcal{X}_J \in \mathcal{C}_{\mathrm{F}}$, where $\mathcal{C}_{\mathrm{F}}$ is defined as in (6) but with $W = K$, such that $\forall\, \mathbf{H} \in \mathcal{A}^a_\Delta,\ \Pr(\mathcal{E}|\mathbf{H}, \mathcal{X}_J) \le \epsilon$. Here, we wish to show that such a code also exists in the smaller set $\mathcal{C}_{\mathrm{E}}$. Instead of reproducing the elaborate set of techniques used in [9], we will just derive the key differences. First, using the fact that $\mathcal{A}^a_\Delta$ is compact along with standard continuity arguments, we obtain $\mathbf{D} = \mathrm{diag}\{1-\delta/2, (1-\delta)\mathbf{1}_{K-1}\}$, where $0 < \delta < 1$, such that

$$\log\left|\mathbf{I} + \mathbf{H}\mathbf{D}\mathbf{H}^\dagger\right| \ge R + \Delta/2,\ \forall\, \mathbf{H} \in \mathcal{A}^a_\Delta. \tag{63}$$

We next generate codewords (or channel inputs) using the density

$$\prod_{j=1}^{J} \frac{1}{\pi^K |\mathbf{D}|} \exp\left(-\mathbf{x}_j^\dagger \mathbf{D}^{-1} \mathbf{x}_j\right), \tag{64}$$

and let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_J]$. The key step required is to show that the probability of the input codeword so generated not lying in (input constraint set) $\mathcal{C}_{\mathrm{E}}$ decays *exponentially* with the blocklength. To do so, we note that $\frac{\mathbf{X}\mathbf{X}^\dagger}{J}$ represents a sample covariance matrix whose expected value is $\mathbf{D}$. Now we need to show that $\Pr\left(\lambda_{\max}\left(\frac{\mathbf{X}\mathbf{X}^\dagger}{J}\right) > 1\right)$ decays exponentially in J. Since $1-\delta/2$ represents the distinct and largest eigenvalue of $\mathbf{D}$, using the results in [23] we can conclude that the $\lambda_{\max}\left(\frac{\mathbf{X}\mathbf{X}^\dagger}{J}\right)$ converges in distribution to a $\mathcal{N}\left(1-\delta/2, \frac{(1-\delta/2)^2}{J}\right)$ random variable as $J \to \infty$. Thus, using the Chernoff upper bound, we can conclude that asymptotically (as $J \to \infty$)

$$\Pr\left(\lambda_{\max}\left(\frac{\mathbf{X}\mathbf{X}^\dagger}{J}\right) > 1\right) \le \frac{1}{2}\exp\left(\frac{-\delta^2 J}{2(2-\delta)^2}\right). \tag{65}$$

The remaining steps remain unchanged from those given in [9]. ■

Note that the restriction on the matrix norm $\|\mathbf{H}\|_\infty \le a$ in the set $\mathcal{A}^a_\Delta$ is an artifact since the proof requires the set to be compact. However, any channel $\mathbf{H}$ with $\|\mathbf{H}\|_\infty > a$ can be written as $\mathbf{H} = \theta\tilde{\mathbf{H}}$, where $\|\tilde{\mathbf{H}}\|_\infty = a$ and the scalar $\theta > 1$. Recall that $\mathcal{X}_J$ is universal over the set $\mathcal{A}^a_\Delta$ so that $\Pr\left(\mathcal{E}\,\middle|\,\tilde{\mathbf{H}}, \mathcal{X}_J\right) \le \epsilon$ because $\tilde{\mathbf{H}} \in \mathcal{A}^a_\Delta$. It is reasonable to assume that

$\Pr\left(\mathcal{E}\left|\mathbf{H},\mathcal{X}_J\right.\right) \leq \Pr\left(\mathcal{E}\left|\tilde{\mathbf{H}},\mathcal{X}_J\right.\right)$, since $\theta > 1$ only ensures a higher received signal power. Thus we can conclude that $\mathcal{X}_J$ is universal over $\mathcal{A}_{\Delta}^{\infty}$, i.e. for any $\mathbf{H}$ with $\log|\mathbf{I}+\mathbf{H}\mathbf{H}^{\dagger}| \geq R+\Delta$, $\Pr\left(\mathcal{E}\left|\mathbf{H},\mathcal{X}_J\right.\right) \leq \epsilon$.

IX. CONCLUSIONS

The connection between the average frame error probability and the outage probability over MIMO fading channels was explored. Three telescoping sets of space-time codes were defined for a given rate and it was shown that FEPs arbitrarily close to the outage probability can be achieved by codes in each set. Further, for a set of codes satisfying constraints stricter than the commonly assumed average or maximum energy constraints, we proved that the outage probability indeed represents the *best achievable* FEP in the large framelength limit. Moreover, the smallest set was also shown to contain universal codes that can communicate reliably over any channel realization for which the mutual information exceeds the information rate of the code.

APPENDICES

A. Stochastic Ordering

The following results are proved in Section 6B of [24]. For any two real-valued random variables X, Y we say $X \geq^{st} Y$ if $\forall\, y \in \mathbb{R}$, $\Pr(X > y) \geq \Pr(Y > y)$.

Then we have

- If $X_1 \geq^{st} Y_1$ and $X_2 \geq^{st} Y_2$, the pair (X_1, X_2) is independent and the pair (Y_1, Y_2) is independent, then

$$X_1 + X_2 \geq^{st} Y_1 + Y_2\ . \tag{66}$$

- Let $\mathcal{CN}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ denote a complex (proper) normal random vector of mean $\boldsymbol{\mu}$ and covariance $\sigma^2\mathbf{I}$. Let $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ be any two vectors s.t. $\|\boldsymbol{\mu}_1\|^2 \geq \|\boldsymbol{\mu}_2\|^2$. Then

$$\|\mathcal{CN}(\boldsymbol{\mu}_1, \sigma^2\mathbf{I})\|^2 \geq^{st} \|\mathcal{CN}(\boldsymbol{\mu}_2, \sigma^2\mathbf{I})\|^2\ . \tag{67}$$

B. Counter-example

We choose a system with $N = 1$ and $K = 2$ and as in [4] we assume $\mathbf{H}$ has i.i.d. $\mathcal{CN}(0,1)$ elements. We consider a sequence of codes (indexed by their length) of rate R, $\{\mathcal{X}_J\}_{J=1}^{\infty}$, $\mathcal{X}_J \in$

$\mathcal{C}_{\rm F}$, where $\mathcal{X}_J = \{\mathbf{X}^j\}_{j=1}^{M_J=\lceil 2^{RJ}\rceil}$ with $\mathbf{X}^j \in \mathbb{C}^{2\times J}$. Define $\zeta^j \triangleq \lambda_{\max}\left(\frac{1}{J}\mathbf{X}^j(\mathbf{X}^j)^\dagger\right)$ and $\psi^j \triangleq \lambda_{\min}\left(\frac{1}{J}\mathbf{X}^j(\mathbf{X}^j)^\dagger\right)$ and suppose that

$$\zeta^j \geq \frac{3W}{4}, \;\; \psi^j \leq \frac{W}{4} \;\; \& \;\; \zeta^j + \psi^j \leq W, \; \forall \;\; 1 \leq j \leq M_J. \tag{68}$$

Note that by the assumption in [4], for *any* fixed $\delta > 0$, we must have

$$\limsup_{J\to\infty} \Pr(\mathbf{Y} \notin S_\delta(\mathbf{H}) | \mathcal{X}_J) = 0, \tag{69}$$

where now $S_\delta(\mathbf{H}) = \left\{\mathbf{Z} \in \mathbb{C}^{1\times J} : \|\mathbf{Z}^\dagger\mathbf{Z}\|_{\rm F}^2 \leq J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2\right\}$. Then letting $\mathbf{X}^j = \mathbf{F}^j\mathbf{\Lambda}^j(\mathbf{G}^j)^\dagger$ denote the SVD of $\mathbf{X}^j$, we have that

$$\begin{aligned}
\Pr(\mathbf{Y} \notin S_\delta(\mathbf{H}) | \mathbf{X}^j) &= \Pr\left(\|\mathbf{H}\mathbf{X}^j + \mathbf{V}\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \middle| \mathbf{X}^j \right) \\
&= \Pr\left(\|(\mathbf{H}\mathbf{F}^j)\mathbf{\Lambda}^j(\mathbf{G}^j)^\dagger + \mathbf{V}\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\mathbf{F}^j\|^2 \middle| \mathbf{X}^j \right) \\
&\overset{(a)}{=} \Pr\left(\|\mathbf{H}\mathbf{\Lambda}^j + \mathbf{V}\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \middle| \mathbf{X}^j \right) \\
&= E\left[\Pr\left(\|\sqrt{|H_{1,1}|^2\zeta_j + |H_{1,2}|^2\psi_j}\mathbf{1}_J + \mathbf{V}\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \middle| \mathbf{H}, \mathbf{X}^j \right) \right]
\end{aligned} \tag{70}$$

where (a) follows since $\mathbf{H}$ and $\mathbf{V}$ are independent of $\mathbf{X}^j$ and unitarily invariant. Using (68) and the stochastic ordering result (67) in (70), we see that

$$\begin{aligned}
&\Pr\left(\left\| \sqrt{|H_{1,1}|^2\zeta_j + |H_{1,2}|^2\psi_j}\;\mathbf{1}_J + \mathbf{V} \right\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \middle| \mathbf{H}, \mathbf{X}^j \right) \\
&\qquad \geq \quad \Pr\left(\|\sqrt{|H_{1,1}|^2 3W/4}\;\mathbf{1}_J + \mathbf{V}\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \middle| \mathbf{H} \right), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \leq j \leq M_J,
\end{aligned} \tag{71}$$

so that

$$\begin{aligned}
&\liminf_{J\to\infty} \Pr\left(\mathbf{Y} \notin S_\delta(\mathbf{H}) \,|\mathcal{X}_J\right) \\
&\qquad \geq \liminf_{J\to\infty} E\left[\Pr\left(\left\| \sqrt{|H_{1,1}|^2 3W/4}\;\mathbf{1}_J + \mathbf{V} \right\|_{\rm F}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\rm F}^2 \,|\mathbf{H} \right) \right]
\end{aligned} \tag{72}$$

Let $\mathcal{A}_\delta(\mathbf{H}) \triangleq \left\{\mathbf{H} : \frac{W}{4}|H_{1,1}|^2 > \frac{W}{2}|H_{1,2}|^2 + 2\delta\right\}$. Then for any $\mathbf{H} \in \mathcal{A}_\delta(\mathbf{H})$, we note that the following inequality holds true.

$$\begin{aligned} & J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\mathrm{F}}^2 - E\left[\|\mathbf{V}\|_{\mathrm{F}}^2\right] - \left\| \sqrt{|H_{1,1}|^2 3W/4}\,\mathbf{1}_J \right\|^2 \\ & \quad = J\left(\delta + |H_{1,2}|^2\frac{W}{2} - |H_{1,1}|^2\frac{W}{4}\right) \\ & \quad < -J\delta. \end{aligned} \tag{73}$$

As a result for any $\mathbf{H} \in \mathcal{A}_\delta(\mathbf{H})$, using the weak law of large numbers we get

$$\lim_{J\to\infty} \Pr\left(\left\| \sqrt{|H_{1,1}|^2 3W/4}\,\mathbf{1}_J + \mathbf{V} \right\|_{\mathrm{F}}^2 > J(\sigma^2+\delta) + \frac{WJ}{2}\|\mathbf{H}\|_{\mathrm{F}}^2 \,\middle|\, \mathbf{H} \right) = 1. \tag{74}$$

Using (74) in (72) along with the dominated convergence theorem finally yields

$$\liminf_{J\to\infty} \Pr(\mathbf{Y} \notin S_\delta(\mathbf{H})|\mathcal{X}_J) \geq \Pr\left(\frac{W}{4}|H_{1,1}|^2 > \frac{W}{2}|H_{1,2}|^2 + 2\delta\right) > 0, \tag{75}$$

which contradicts the assumption in (69).

REFERENCES

[1] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

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

[3] G. Caire and G. Colavolpe, "On space-time coding for quasi-static multiple-antenna channels," submitted to *IEEE Trans. Inform. Theory*, vol. 49, no. 6, pp. 1400–1416, June 2003.

[4] M. Fozunbal, S. W. McLaughlin, and R. W. Schafer, "On performance limits of space-time codes: a sphere-packing bound approach," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2681–2687, Oct. 2003.

[5] N. Prasad and M. K. Varanasi, "Bounds on error probabilities for the block fading channel," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Monterey, CA, Nov. 2004, (invited).

[6] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Enginering*, John Wiley & Sons, 1965.

[8] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999, Originally a Bell Laboratories, Lucent Technologies, Technical Report, Oct. 1995.

[9] W. L. Root and P. P. Varaiya, "Capacity of classes of gaussian channels," *Ann. Math. Stat.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[10] H. Boche and E. A. Jorsweick, "Outage probability of multiple antenna systems: Optimal transmission and impact of correlation," in *Proc. Intl. Zurich Seminar on Communications*, Zurich, Switzerland, Feb. 2004, pp. 116–119.

[11] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic space-time architectures," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 303–320, Apr. 2003.

[12] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[13] C. Kose and R. Wesel, “Universal space-time trellis codes,” *IEEE Trans. Inform. Theory*, vol. 46, no. 10, pp. 2717–2727, Oct. 2003.

[14] S. Tavildar and P. Viswanath, “Permutation codes: Achieving the diversity-multiplexing tradeoff,” in *Proc. IEEE Intl. Symposium on Information Theory*, Chicago, IL, June 2004, p. 352.

[15] S. Tavildar and P. Viswanath, “Approximately universal codes over slow fading channels,” submitted to *IEEE Trans. Inform. Theory*, 2005.

[16] H. Royden, *Real Analysis*, Prentice Hall, Englewood Cliffs, NJ, third edition, 1988.

[17] R. G. Gallager, *Information Theory and Reliable Communications*, John Wiley, New York, 1968.

[18] T. Guess and M. K. Varanasi, “Error exponents for maximum-likelihood and successive decoders for the gaussian cdma channel,” *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1683–1690, July 2000.

[19] J. W. Lamperti, *Probability*, John Wiley & Sons, 2nd edition, 1996.

[20] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables*, Marcel Dekker, Inc., New York, 1992.

[21] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 3rd edition, 1995.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, U.K., 2004.

[23] T. Anderson, “Asymptotic theory for principle component analysis,” *Ann. Math. Stat.*, vol. 34, pp. 122–148, 1963.

[24] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multi-antenna systems on flat fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.