Dispersion of the Coherent MIMO Block-Fading Channel

Austin Collins and Yury Polyanskiy

***Abstract*—In this paper we consider a channel model that is often used to describe the mobile wireless scenario: multiple-antenna additive white Gaussian noise channels subject to random (fading) gain with full channel state information at the receiver. Dynamics of the fading process are approximated by a piecewise-constant process (frequency non-selective isotropic block fading). This work addresses the finite blocklength fundamental limits of this channel model. Specifically, we give a formula for the channel dispersion – a quantity governing the delay required to achieve capacity – and present achievability and (partial) converse bounds. Multiplicative nature of the fading disturbance leads to a number of interesting technical difficulties that required us to enhance traditional methods for finding channel dispersion. Knowledge of channel dispersion opens the possibility for studying the impact of channel dynamics, antenna selection rules, etc on the communication rate.**

I. Introduction

Given a noisy communication channel, the maximal cardinality of a codebook of blocklength n which can be decoded with block error probability no greater than ϵ is denoted as $M^*(n,\epsilon)$. Evaluation of this function – the fundamental performance limit of block coding – is computationally impossible for most channels of interest. To resolve this difficulty, [1] proposed a closed-form normal approximation, based on the asymptotic expansion:

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,, \qquad (1)$$

where capacity C and dispersion V are two intrinsic characterstics of the channel and $Q^{-1}(\epsilon)$ is the inverse of the Q-function[1]. One immediate consequence of the normal approximation is an estimate for the minimal blocklength (delay) required to achieve a given fraction η of channel capacity:

$$n \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1-\eta}\right)^2 \frac{V}{C^2}\,. \qquad (2)$$

Asymptotic expansions such as (1) are rooted in the central-limit theorem and have been known classically for discrete memoryless channels [2], [3] and later extended in a wide variety of directions; see [4] for a survey.

Motivated by a recent surge of orthogonal frequency division (OFDM) technology, this paper focuses on the frequency-nonselective coherent complex block fading discrete-time channel with multiple transmit and receive antennas (MIMO) (See [5, Section II] for background on this model). Formally, let $n_t \geq 1$ be the number of transmit antennas, $n_r \geq 1$ be the number of receive antennas, and $T \geq 1$ be the coherence time of the channel. The input-output relation at block j (spanning time instants $(j-1)T+1$ to jT) with $j = 1,\ldots,n$ is given by

$$\mathbb{Y}_j = \mathbb{H}_j \mathbb{X}_j + \mathbb{Z}_j\,, \qquad (3)$$

where $\{\mathbb{H}_j, j = 1,\ldots\}$ is a $n_r \times n_t$ matrix-valued random fading process, $\mathbb{X}_j$ is a $n_t \times T$ matrix channel input, $\mathbb{Z}_j$ is a $n_r \times T$ Gaussian random matrix with independent circularly symmetric entries of zero mean and unit variance, and $\mathbb{Y}_j$ is the $n_r \times T$ matrix-valued channel output. The process $\mathbb{H}_j$ is assumed to be i.i.d. with isotropic distribution $P_{\mathbb{H}}$, i.e. for any unitary matrices $U \in \mathbb{C}^{n_r \times n_r}$ and $V \in \mathbb{C}^{n_t \times n_r}$, both $U\mathbb{H}$ and $\mathbb{H}V$ are equal in distribution to $\mathbb{H}$. The fading also satisfies the normalization condition

$$\mathbb{E}\left[\|\mathbb{H}\|_F^2\right] = 1\,. \qquad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Note that due to merging channel inputs at time instants $1,\ldots,T$ into one matrix-input, the block-fading channel becomes memoryless. We assume coherent demodulation so that the channel state information (CSI) $\mathbb{H}_j$ is fully known to the receiver (CSIR).

An $(nT, M, \epsilon, P)_{CSIR}$ code of blocklength nT, probability of error ϵ and power-constraint P is a pair of maps: the encoder $f : [M] \to (\mathbb{C}^{n_t \times T})^n$ and the decoder $g : (\mathbb{C}^{n_r \times T})^n \times (\mathbb{C}^{n_r \times n_t})^n \to [M]$ such that

$$\mathbb{P}[W \neq \hat{W}] \leq \epsilon\,. \qquad (5)$$

and

$$\sum_{j=1}^n \|\mathbb{X}_j\|_F^2 \leq nTP \qquad \mathbb{P}\text{-a.s.}\,,$$

on the probability space

$$W \to \mathbb{X}^n \to (\mathbb{Y}^n, \mathbb{H}^n) \to \hat{W}\,,$$

where W is uniform on $[M]$, $\mathbb{X}^n = f(W)$, $\mathbb{X}^n \to (\mathbb{Y}^n, \mathbb{H}^n)$ is as described in (3) and $\hat{W} = g(\mathbb{Y}^n, \mathbb{H}^n)$.

Under the isotropy assumption on P_H, the capacity C of this channel is given by [6]

$$C(P) = \mathbb{E}\left[\log\det\left(I_{n_r} + \frac{P}{n_t}\mathbb{H}\mathbb{H}^\dagger\right)\right] \qquad (6)$$

where I_n denotes the $n \times n$ identity matrix.

The goal of this line of work is to characterize dispersion of the present channel. Since the channel is memoryless it is natural to expect, given the results in [1], [7], that dispersion (for $\epsilon < 1/2$) is given by

$$V_{min} = \inf_{P_X : I(X;Y|H)=C} \frac{1}{T}\mathrm{Var}[i(X;Y,H)|X] \qquad (7)$$

Authors are with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA. e-mail: {austinc,yp}@mit.edu.

This material is based upon work supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

[1] As usual, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$.

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

where the information density is denoted

$$i(x; y, h) \triangleq \log \frac{dP_{\mathbb{Y},\mathbb{H}|\mathbb{X}=x}}{dP^*_{\mathbb{Y},\mathbb{H}}}(y, h) \tag{8}$$

and $P^*_{\mathbb{Y},\mathbb{H}}$ is a capacity-achieving output distribution. In this work, we justify (7) as the actual (operational) dispersion, appearing in the expansion of $\log M^*(n, \epsilon)$. A general achievability bound is given, along with a partial converse result that assumes the supreumum norm of all codewords is bounded by $\|x^n\|_\infty = o(n^{1/4})$. Without this constraint, information density may not be asymptotically normal and the general (unconstrained) converse is still open. Both the achievability and partial converse proofs require novel steps as compared to techniques invoked in previous works on channel dispersion and finite blocklength information theory.

In the MISO case ($n_r = 1$), the solution to (7) is non-trivial, because the class of input distributions that achieve capacity is rich. This was analyzed by us in [8], where it was shown that full rate orthogonal designs are the solution to (7) in dimensions where they exist (e.g. Alamouti code for $n_t = T = 2$). In this work, we focus on the case where $n_r \geq 2$. In this case, the capacity achieving input distribution is unique, so the minimization is trivial and we are able to give a closed-form expression for (7).

Before proceeding to our results, we mention recent literature on dispersion of wireless channels. Single antenna channel dispersion is computed in [7] for a coherent channel subject to stationary fading process. In [9] finite-blocklength effects are explored for the non-coherent (i.e. no CSI) block fading setup. Quasi-static fading channels in the general MIMO setting have been thoroughly investigated in [10], showing that the expansion (1) changes dramatically (in particular the channel dispersion term becomes zero); see also [11] for evaluation of the bounds. Coherent quasi-static channel has been studied in the limit of infinitely many antennas in [12] appealing to concentration properties of random matrices. Dispersion for lattices (infinite constellations) in fading channels has been investigated in a sequence of works, see [13] and references. The regime of low spectral efficiency (or minimum energy-per-bit) for general MIMO channel in both the coherent and non-coherent case was studied in [14]. [15] investigates how power control (i.e. power constraint averaged over the codebook) affects fundamental limits under the quasi-static fading channel with perfect CSIRT.

All proofs in this paper can be found in [16]. The numerical tool used to compute the achievability bounds, dispersion, and normal approximation are available online [17].

II. Main Results

Our first result is a coding theorem giving the achievability and partial converse for the MIMO coherent fading channel.

Theorem 1. *For the MIMO-BF channel, there exists an $(nT, M, \epsilon, P)_{CSIR}$ maximal probability of error code with $0 < \epsilon < 1/2$ and*

$$\log M \geq nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + o(\sqrt{n}) \tag{9}$$

Furthermore, for any $\delta_n \to 0$ there exists $\delta'_n \to 0$ so that every $(nT, M, \epsilon, P)_{CSIR}$ code with extra constraint that $\|x^n\|_\infty \leq \delta_n n^{1/4}$, must satisfy

$$\log M \leq nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + \delta'_n\sqrt{n} \tag{10}$$

where $C(P)$ is the capacity given in (6), and

$$V(P) = \inf_{P_{\mathbb{X}}: I(\mathbb{X};\mathbb{Y}|\mathbb{H}) = C} \frac{1}{T}\mathrm{Var}[i(\mathbb{X}; \mathbb{Y}, \mathbb{H})|\mathbb{X}] \tag{11}$$

and $i(x; y, h)$ is the information density, given in (19) below.

In the case when $n_r \geq 2$, the distribution achieving capacity is unique. In this case, we can compute the conditional variance (11) in closed form. The expression is given in the following theorem.

Theorem 2. *When $n_r \geq 2$,*

$$\begin{aligned} V(P) = {} & T\mathrm{Var}\left(\sum_{i=1}^{n_{\min}} C_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right) \\ & + n_{\min}\mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda^2\right)\right] \\ & + n_{\min}\left(\frac{P}{n_t}\right)^2 \chi_1 \\ & - \left(\frac{P}{n_t}\right)^2 \chi_2 \frac{n_{\min}^2}{n_t} \end{aligned} \tag{12}$$

where $\Lambda_i^2, i = 1, \ldots n_{\min}$ are eigenvalues of $\mathbb{H}\mathbb{H}^\dagger$, Λ^2 is the marginal distribution a single eigenvalue of $\mathbb{H}\mathbb{H}^\dagger$, and

$$C_{AWGN}(P) = \log(1 + P) \tag{13}$$

$$V_{AWGN} = \log^2(e)\left(1 - \frac{1}{(1+P)^2}\right) \tag{14}$$

$$\chi_1 = \log^2(e)\mathbb{E}\left[\left(\frac{\Lambda^2}{1 + \frac{P}{n_t}\Lambda^2}\right)^2\right] \tag{15}$$

$$\chi_2 = \log^2(e)\mathbb{E}^2\left[\frac{\Lambda^2}{1 + \frac{P}{n_t}\Lambda^2}\right] \tag{16}$$

Note here that all eigenvalues $\Lambda_1^2, \ldots, \Lambda_{n_{\min}}^2$ of $\mathbb{H}\mathbb{H}^\dagger$ have the same marginal distribution by the assumption that $\mathbb{H}$ is isotropic, so Λ^2 is well defined.

In Section II-A we review some relevant properties of the coherent block fading channel. Then in Sections II-B and II-C we outline the strategies to prove the achievability and converse bounds, respectively. In Section II-D we discuss the implications of the dispersion results on communicating over fading channels.

A. Preliminaries

The capacity of this channel (6) was first proved by Telatar [6]. Telatar also showed that the input distribution with i.i.d. $\mathcal{CN}(0, P/n_t)$ (circularly symmetric Gaussian) entries achieves capacity. In this work, the *capacity achieving output distribution* (caod) will be of interest, i.e. the distribution $P_{\mathbb{Y}\mathbb{H}}$ induced by the capacity achieving input distribution $P^*_{\mathbb{X}}$ through the channel. Here, $P^*_{\mathbb{Y}\mathbb{H}} = P_{\mathbb{H}}P^*_{\mathbb{Y}|\mathbb{H}}$ where $P_{\mathbb{H}}$ is distribution of the fading process, and

$$P^*_{\mathbb{Y}|\mathbb{H}} \sim \mathcal{CN}\left(0, I_{n_r} + \frac{P}{n_t}\mathbb{H}\mathbb{H}^\dagger\right) \tag{17}$$

In [8], the authors show that when $n_r = 1$, the distribution that achieves capacity is not unique and not necessarily Gaussian. However, as soon as $n_r \geq 2$, the following proposition states the Telatar's distribution is a unique maximizer of capacity:

Proposition 3. *When $n_r \geq 2$, the capacity achieving input distribution is uniquely $\mathbb{X} \in \mathbb{C}^{n_t \times T}$ where each entry X_{ij} has an i.i.d. $\mathcal{N}(0, P/n_t)$ distribution.*

Proof sketch: Any capacity achieving input distribution induces the unique output distribution (17). This implies that for any $h \in \mathbb{C}^{n_r \times n_t}$, and $X_i, X_j \in \mathbb{C}^{n_t \times 1}$, $i \neq j$ being any two columns of $\mathbb{X}$, we have

$$h\mathbb{E}[\mathbb{X}_i \mathbb{X}_j^\dagger]h^\dagger = 0_{n_r}. \tag{18}$$

where O_n denotes the $n \times n$ all zero matrix. When $n_r > 1$, (18) can hold only if $\mathbb{E}[\mathbb{X}_i \mathbb{X}_j^\dagger]$ is the all zero matrix. Along with other Gaussian constraints, this forces $\mathbb{X}$ to be unique. ■

A fundamental quantity needed to prove finite blocklength bounds is the information density, as defined in (8). Some algebra yields the explicit expression for the (single-letter) information density as follows:

$$\begin{aligned} i(x; y, h) &= T \log\det\left(I_{n_r} + \frac{P}{n_t} hh^\dagger\right) \\ &+ \sum_{j=1}^{n_{\min}} \frac{\|\Lambda_j w_j\|^2 + \Lambda_j \langle w_j, \tilde{z}_j\rangle + \bar{\Lambda}_j \langle \tilde{z}_j, w_j\rangle - \frac{P}{n_t}|\Lambda_j|^2 \|z_j\|^2}{1 + \frac{P}{n_t}|\Lambda_j|^2} \end{aligned} \tag{19}$$

where $\bar{a}$ is the complex conjugate of $a \in \mathbb{C}$, and

1) $n_{\min} = \min(n_t, n_r)$.
2) $h = UDV^\dagger$ is the SVD of h, where $D \in \mathbb{C}^{n_r \times n_t}$ with $\Lambda_1, \ldots, \Lambda_{n_{\min}}$ on the principal diagonal and zeros elsewhere.
3) $w = V^\dagger x$, and w_j denotes the j-th row of w.
4) $\tilde{z} = U^\dagger z$, and $\tilde{z}_j$, z_j denote the j-th row of $\tilde{z}$, z, respectively.

Theorem 1 justifies the quantity $\frac{1}{T}\mathrm{Var}[i(\mathbb{X}; \mathbb{Y}, \mathbb{H})|\mathbb{X}]$ as the quantity to compute for the operational dispersion, and Theorem 2 gives the closed form expression for this conditional variance.

B. Achievability

The proof utilizes the $\kappa\beta$ bound [1], which states that for a given a channel $P_{\mathbb{Y}|\mathbb{X}}$ with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$, for any distribution Q_Y on $\mathcal{B}$, and ϵ, τ such that $0 < \tau < \epsilon < 1$, there exists an (M, ϵ) code satisfying

$$M \geq \frac{\kappa_\tau(F, Q_Y)}{\sup_{x \in F} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)} \tag{20}$$

where F is any set in the input space.

The notation $\beta_\alpha(P, Q)$ denotes the minimal error in a binary hypothesis test between distribution P and Q given that the test chooses P when P is the true distribution with at least probability α. $\kappa_\tau(F, Q)$ for set F and distribution Q denotes the minimum error in a composite hypothesis test, which decides between the set of distributions $\{P_{Y|X=x}\}_{x \in F}$ and Q. For more background on these definitions, see [1].

The "art" of applying the $\kappa\beta$ bound is in choosing F and $Q_{\mathbb{Y}\mathbb{H}}$ appropriately. In this case, and as is common, $Q_{\mathbb{Y}\mathbb{H}}$ is chosen to be the capacity achieving output distribution (17). To motivate the choice of the set F, the following lower bound on $\kappa_\tau(F, Q_{\mathbb{Y}\mathbb{H}})$ can be shown

$$\begin{aligned} &\kappa_\tau(F, Q_{\mathbb{Y}^n\mathbb{H}^n}) \\ &\quad \geq \exp\left(-\frac{D(P_{\tilde{\mathbb{X}}^n} \circ P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n} || Q_{\mathbb{Y}^n\mathbb{H}^n}) + \log 2}{\tau P_{\tilde{\mathbb{X}}^n}[F]}\right) \end{aligned} \tag{21}$$

where the notation $P_X \circ P_{Y|X}$ denotes the output distribution induced through the channel $P_{Y|X}$ by input P_X, and $P_{\tilde{\mathbb{X}}^n}$ is any distribution such that $P_{\tilde{\mathbb{X}}^n}[F] > 0$. The key reason why (21) is a useful lower bound is the following lemma.

Lemma 4. *Let $\mathbb{X}$ be any capacity achieving input distribution for the channel, and $\mathbb{X}^n = [\mathbb{X}_1 \cdots \mathbb{X}_n]$. Let*

$$P_{\tilde{\mathbb{X}}^n} \sim \frac{\mathbb{X}^n}{\|\mathbb{X}^n\|_F} \sqrt{nTP} \tag{22}$$

and $Q_{Y^n H^n} = \prod_{i=1}^n P_{\mathbb{Y}\mathbb{H}}^*$, where $P_{\mathbb{Y}\mathbb{H}}^*$ is the (unique) capacity achieving output distribution as described in (17)*. Then there exists a constant $K > 0$ such that for all $n = 1, 2, \ldots$,*

$$D(P_{\tilde{\mathbb{X}}^n} \circ P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n} || Q_{\mathbb{Y}^n\mathbb{H}^n}) \leq K \tag{23}$$

(A similar result was shown for the AWGN channel by MolavianJazi et al [18, Proposition 2].) This Lemma can be interpreted as saying: the output distribution induced by any caid normalized to lie on the manifold $\{x^n \in \mathbb{C}^{nn_t \times T} : \|x^n\|_F^2 = nTP\}$ is similar (in the sense of divergence) to the caod. This choice of $P_{\tilde{\mathbb{X}}^n}$ is good for the numerator in the bound (21), and this motivates the choice of the set F, i.e. we need $P_{\tilde{\mathbb{X}}^n}[F]$ to be bounded away from zero for all n. The following Lemma gives conditions for this to hold.

Lemma 5. *For $\mathbb{X}$ caid, $\tilde{\mathbb{X}}^n$ distributed as $\frac{\mathbb{X}^n}{\|\mathbb{X}^n\|_F}\sqrt{nTP}$, we have*

$$\mathbb{P}\left[\left|\frac{1}{T}V_n(\tilde{\mathbb{X}}^n) - \frac{1}{T}\mathbb{E}[V_1(\mathbb{X})]\right| \geq \delta\right] \to 0 \quad \text{as} \quad n \to \infty \tag{24}$$

where

$$V_n(x^n) = \frac{1}{n}\sum_{j=1}^n \mathrm{Var}[i(\mathbb{X}_j; \mathbb{Y}_j, \mathbb{H}_j)|\mathbb{X}_j = x_j] \tag{25}$$

In this notation, $\mathbb{E}[V_1(\mathbb{X})] = \mathrm{Var}[i(\mathbb{X}; \mathbb{Y}, \mathbb{H})|\mathbb{X}]$.

The function $V_n(x^n)$ is the *empirical conditional variance*. With this in mind, choose the set F to be

$$F = \{x^n : \|x^n\|_F^2 = nTP\} \cap \left\{x^n : \frac{1}{T}V_n(x^n) \leq V + \delta\right\} \tag{26}$$

Lemma 5 guarantees that F chosen as (26) satisfies $P_{\tilde{\mathbb{X}}^n}[F] \to 1$. With these choices of F and $Q_{\mathbb{Y}^n\mathbb{H}^n}$, $\kappa_\tau(F, Q_{\mathbb{Y}^n\mathbb{H}^n})$ is lower bounded by a constant. The denominator of (20) can be expanded in the usual way using the Berry-Esseen theorem, giving the expression in the achievability bound.

C. Converse

The converse utilizes the maximum probability of error meta-converse bound: any (n, M, ϵ) maximum probability of error code satisfies

$$\frac{1}{M} \geq \inf_{x^n \in F} \beta_{1-\epsilon}\left(P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}, Q_{\mathbb{Y}^n\mathbb{H}^n}\right) \quad (27)$$

where $F = \{x^n \in \mathbb{R}^{nn_t \times T} : \|x^n\|_F^2 = nTP\}$ is the constraint set for the channel, and $Q_{\mathbb{Y}^n\mathbb{H}^n}$ is any distribution on the output space. Again, $Q_{\mathbb{Y}^n\mathbb{H}^n}$ is chosen to be the i.i.d. n-fold product of capacity achieving output distributions (17). There are a few challenges that emerge: first, to apply central limit theorem based theorems such as the Berry Esseen theorem, the distribution of the log likelihood ratio

$$\log \frac{P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}}{Q_{\mathbb{Y}^n\mathbb{H}^n}}(\mathbb{Y}^n, \mathbb{H}^n) \qquad |\mathbb{X}^n = x^n \quad (28)$$

must be asymptotically normal. For the MIMO fading channel, this is not true for all values of $x^n \in F$. Specifically, (28) is asymptotically normal when x^n is not too "peaky", i.e. when $\|x^n\|_\infty = o(n^{1/4})$, which is the primary reason for imposing this constraint. Without the constraint, for example the codeword where one entry has value $\sqrt{nTP}$ and all the other entries are zero clearly does not result in the information density being asymptotically normal.

The second difficulty arises as follows: when $\|x^n\|_\infty = o(n^{1/4})$, we have the expansion

$$\begin{aligned} -\log&\beta_{1-\epsilon}\left(P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}, Q_{\mathbb{Y}^n\mathbb{H}^n}\right) \\ &= \mathbb{E}[i(x^n;\mathbb{Y}^n,\mathbb{H}^n)] \\ &\quad - \sqrt{\mathrm{Var}(i(x^n;\mathbb{Y}^n,\mathbb{H}^n))}Q^{-1}(\epsilon) + o(\sqrt{n}) \quad (29) \end{aligned}$$

Here, the $O(n)$ term is constant over all $x^n \in F$. The $O(\sqrt{n})$ term is not constant over F, and furthermore, "smooth" x^n's, e.g. where all entries are $\sqrt{P/n_t}$, causes $\mathrm{Var}(i(x^n;\mathbb{Y}^n,\mathbb{H}^n))$ to be smaller than the conditional variance (11), and hence not tight with the achievability bound. To remedy this, we expurgate these "smooth" codewords using the following technique. The intuition is that although some x^n's give an unusually small dispersion, the set of such x^n's is too small to support a capacity-achieving code. Given a codebook $\mathcal{C}$, fix $\delta > 0$, and split it into two pieces

$$\mathcal{C}_l \triangleq \mathcal{C} \cap \{x^n : V_n(x^n) \leq n(V-\delta)\} \quad (30)$$

$$\mathcal{C}_u \triangleq \mathcal{C} \cap \{x^n : V_n(x^n) > n(V-\delta)\} \quad (31)$$

where $V_n(x^n)$ is given in (25). A converse is shown for each of the pieces separately: if M_l, M_u are the number of codewords in $\mathcal{C}_l, \mathcal{C}_u$, respectively, then

$$\log M \leq \log\max(M_u, M_l) + \log 2 \quad (32)$$

showing an upper bound on $\log M_u$ is simple: using the meta converse and expansion (29), we conclude that

$$\log M_u \leq nTC - \sqrt{nT(V-\delta)}Q^{-1}(\epsilon) + o(\sqrt{n}) \quad (33)$$

The $V-\delta$ term follows immediately from the form of the codebook $\mathcal{C}_u$. The portion $\mathcal{C}_l$ requires more work. Define $F_l \triangleq \{x^n : V(x^n) \leq V - \delta\}$. We prove the following

Lemma 6. *For any $\delta > 0$ and any $(nT, M, \epsilon, P)_{CSIR}$ maximum probability of error code (for the MIMO-BF channel) with codewords in $F_l \triangleq \{x^n : V_n(x^n) \leq V - \delta\}$, there exists $\tau > 0$ such that*

$$\log M \leq n(C - \tau) + O(1) \quad (34)$$

where C is the capacity (6).

Remark 1. This Lemma gives the strong converse for the fading channel with the additional codebook constraint $V_n(x^n) \leq V - \delta$, and helps bound $\log M_l$.

Proof sketch: We apply the following basic bound, for any $\gamma_n > 0$,

$$\begin{aligned} \frac{1}{M} &\geq \inf_{x \in F_l} \beta_{1-\epsilon}(P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}, Q_{\mathbb{Y}^n\mathbb{H}^n}) \quad (35) \\ &\geq \inf_{x \in F_l} \frac{1}{\gamma_n}\Bigg(1 - \epsilon - \\ &\qquad \mathbb{P}\left[\log \frac{P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}}{Q_{\mathbb{Y}^n\mathbb{H}^n}}(\mathbb{Y}^n,\mathbb{H}^n) \geq \log\gamma_n\right]\Bigg) \quad (36) \end{aligned}$$

On $\mathcal{C}_l$, we choose $Q_{\mathbb{Y}^n\mathbb{H}^n}$ to be the capacity achieving output distribution *of the channel with the additional constraint*,

$$C' = \frac{1}{T}\sup_{\substack{P_X : \mathbb{E}[\|\mathbb{X}\|_F^2] \leq TP \\ \mathbb{E}[V_1(\mathbb{X})] \leq V - \delta}} I(\mathbb{X};\mathbb{Y}|\mathbb{H}) \quad (37)$$

For this modified capacity problem, we do not have an expression for the capacity or the capacity achieving input or output distributions. However, it can be shown that supremum is attained. General results on capacity also imply that there exists a unique capacity achieving output distribution $Q_{\mathbb{Y}\mathbb{H}}$ for (37). Knowing that $Q_{\mathbb{Y}\mathbb{H}}$ is induced through the channel implies that the distribution should be sufficiently well-behaved, namely:

Lemma 7. *Let $P_{\mathbb{Y}^n\mathbb{H}^n}$ be any distribution induced through the MIMO-BF channel by some $P_{\mathbb{X}^n}$ satisfying $\mathbb{E}[\|\mathbb{X}\|_F^2] = nTP$. Then for any x^n satisfying $\|x^n\|_F^2 = nTP$, there exists a constant $K > 0$ such that for all n,*

$$\mathrm{Var}\left[\log P_{\mathbb{Y}^n\mathbb{H}^n}(\mathbb{Y}^n,\mathbb{H}^n)|\mathbb{H}^n, \mathbb{X}^n = x^n\right] < nK \quad (38)$$

Lemma 7 is proved via an application of the Poincare inequality for the variance of a function of a Gaussian random variable (25), similar to [19, Theorem 18 and (106)] for the non-fading AWGN channel.

We apply Lemma 7 to the bound (36). Choose $\log\gamma_n = n(C + C')/2$, and using Chebyshev's inequality and some algebra, we obtain

$$\mathbb{P}\left[\log \frac{P_{\mathbb{Y}^n\mathbb{H}^n|\mathbb{X}^n=x^n}}{Q_{\mathbb{Y}^n\mathbb{H}^n}}(\mathbb{Y}^n,\mathbb{H}^n) \geq \log\gamma_n\right] \leq \frac{4n_rT + 2K}{n(C - C')^2} \quad (39)$$

Applying this result to the bound (36), we see that for large enough n,

$$\begin{aligned} \log M_l &\leq \log\gamma_n - \log(1 - \epsilon - O(1/n)) \quad (40) \\ &= n\frac{C + C'}{2} + O(1) \quad (41) \end{aligned}$$

Then $\tau = (C - C')/2$ concludes the argument. ■

D. Discussion

First we must assess the accuracy of the normal approximation (1) by comparing it against numerical bounds. It turns out that the $\kappa\beta$ bound we used for analysis is not the tightest

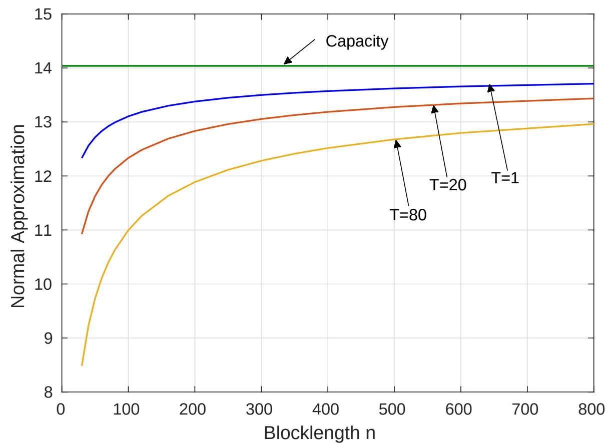

Fig. 1. The normal approximation for varying coherent times, with $n_t = n_r = T = 4$, $P = 20$, and $\epsilon = 10^{-3}$ for Rayleigh fading.

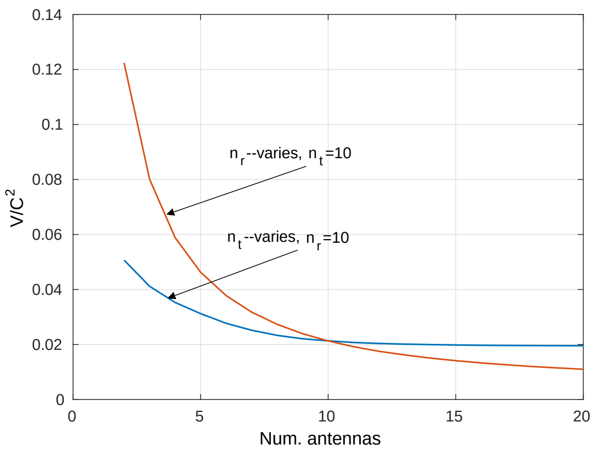

Fig. 2. Dependence of V/C^2 on n_r and n_t for $P = 20$ and $T = 10$. The two curves show one parameter (n_r or n_t) fixed at 10, and the other varying from 2 to 20. The fading distribution is Rayleigh.

achievability. A better bound, a so called $\beta\beta$-bound [20], is computed in [20, Figure 2]. As shown there, (1) gives a reasonable approximation of the achievability bound.

Now we turn to the communication insights of (1) together with the expression for dispersion (12). The dependence on the coherence time clearly has an affine relationship with dispersion. Figure 1 shows the normal approximation for different values of coherence time. Note that this plot was made with H has i.i.d. $\mathcal{CN}(0,1)$ entries, i.e. for Rayleigh fading, to compare our results to canonical values. This is different than the normalization (4), but does not effect any of the results.

The relationship of (12) relative to the number of transmit and receive antennas is less obvious, since the distribution of Λ depends on the number of antennas. Notice that the expression (12) is not symmetric in n_t and n_r. This is shown in Figure 2, which plots the normalized dispersion V/C^2 as a function of n_t and n_r. The blue curve shows that as we increase the number of transmit antennas above 10 (the number of receive antennas), the normalized dispersion is approximately constant. Note that this is a statement about the dispersion, different from the results that capacity scales linearly with the minimum number of antennas.

REFERENCES

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, Berkeley, CA, USA, 1961, pp. 211–252.

[3] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Trans. 3d Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

[4] Y. Polyanskiy and S. Verdú, "Finite blocklength methods in information theory (tutorial)," in *2013 IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/ISIT13_tutorial.pdf

[5] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

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

[7] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2959–2963.

[8] A. Collins and Y. Polyanskiy, "Orthogonal designs optimize achievable dispersion for coherent miso channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, Hawaii, July 2014, pp. 2524–2582.

[9] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 577–581.

[10] ——, "Quasi-static multiple-antenna fading channels at finite block-length," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.

[11] E. MolavianJazi and J. N. Laneman, "On the second-order coding rate of non-ergodic fading channels," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, USA, Oct. 2013.

[12] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO quasi-static Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6591–6622, Dec. 2015.

[13] S. Vituri, "Dispersion analysis of infinite constellations in ergodic fading channels," *arXiv preprint arXiv:1309.4638*, vol. M.S. Thesis, Tel Aviv University, 2013.

[14] W. Yang, G. Durisi, and Y. Polyanskiy, "Minimum energy to send k bits over multiple-antenna fading channels," *arXiv preprint*, July 2015. [Online]. Available: http://arxiv.org/pdf/1507.03843v1.pdf

[15] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598 – 4615, Sept. 2015.

[16] A. Collins and Y. Polyanskiy, "Coherent MIMO block-fading channel at finite blocklength," *preprint*, Jan. 2016. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/bfmimo_journal.pdf

[17] Spectre, "SPECTRE: Short packet communication toolbox," 2015, GitHub repository. [Online]. Available: https://github.com/yp-mit/spectre

[18] E. MolavianJazi and J. N. Laneman, "A finite-blocklength perspective on gaussian multi-access channels," *IEEE Trans. Inf. Theory*, Sep. 2013, submitted. [Online]. Available: http://arxiv.org/abs/1309.2343

[19] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability (extended)," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, Jan. 2014.

[20] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and V. Poor, "A beta-beta achievability bound with applications," *2016 IEEE Int. Symp. Inf. Theory (ISIT)*, Jan. 2016, submitted.