Orthogonal Designs Optimize Achievable Dispersion for Coherent MISO Channels

Austin Collins and Yury Polyanskiy

***Abstract*—This work addresses the question of finite blocklength fundamental limits of coherently demodulated multi-antenna channels, subject to frequency non-selective isotropic fading. Specifically we present achievability bound for the channel dispersion – a quantity known to determine the delay required to achieve capacity. It is shown that a commonly used isotropic Gaussian input, which is only one of many possible capacity achieving distributions, is suboptimal. Optimal inputs minimizing channel dispersion turn out to include a family of modulation techniques known as orthogonal designs (in particular, Alamouti's scheme). For 8 transmit antennas numerical evaluation shows that up to 40% of additional penalty in delay is incurred by using isotropic codewords (compared to dispersion-optimal architecture exploiting transmit diversity).**

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 alas computationally impossible for most channels of interest. As a resolution of 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 characteristics 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 real block fading discrete-time channel with multiple transmit antennas and a single receive antenna (MISO). (See [5, Section II] for background on this model.) Formally, let $n_t \geq 1$ be the number of transmit 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 $j = 1,\dots,n$ is given by

$$Y_j = H_j X_j + Z_j\,, \qquad (3)$$

where $\{H_j, j = 1,\dots\}$ is a $1\times n_t$ vector-valued random fading process, X_j is a $n_t \times T$ matrix channel input, Z_j is a $1\times T$ Gaussian random vector with independent entries of variance 1, and Y_j is the $1\times T$ vector-valued channel output. The process H_j is assumed to be i.i.d. with isotropic distribution P_H satisfying

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

Rayleigh fading is a special case of this assumption where H_j is an i.i.d. Gaussian random vector. Note that because of merging channel inputs at time instants $1,\dots,T$ into one matrix-input, the block-fading channel becomes memoryless. We assume coherent demodulation so that the channel state information 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 $f : [M] \to (\mathbb{R}^{n_t\times T})^n$ and the decoder $g : (\mathbb{R}^{1\times T})^n \times (\mathbb{R}^{1\times n_t})^n \to [M]$ such that

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

and

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

($\|A\|_F^2 = \sum_{ij}|a_{ij}|^2$ is the Frobenius norm of the matrix) on the probability space

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

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

Under the isotropy assumption on P_H, the capacity C appearing in (1) of this channel is given by [6]

$$C(P) = \mathbb{E}\left[C_{AWGN}\left(\frac{P}{n_t}\|H\|^2\right)\right]\,, \qquad (6)$$

where $C_{AWGN}(P) = \frac{1}{2}\log(1+P)$ is the capacity of the additive white Gaussian noise (AWGN) channel with SNR P.

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

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

(for $\epsilon < 1/2$) is given by

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

where we denoted information density

$$i(x;y,h) \triangleq \log \frac{dP_{Y,H|X=x}}{dP_{Y,H}}(y,h)$$

and $P_{X,Y,H} = P_X P_{Y,H|X}$. Justification of (7) as the actual (operational) dispersion, appearing in the expansion of $\log M^*(n,\epsilon)$ is by no means trivial. The present note focuses on the achievability part and on solving minimization (7). The converse will be presented elsewhere.

Because the set of capacity achieving input distributions (c.a.i.d.) is large, minimization of (7) is not trivial. It will be shown below that Hurwitz-Radon families of matrices (introduced into communication in the form of orthogonal designs by Tarokh et al in [8]), and in particular Alamouti's scheme [9], minimize dispersion. This shows a somewhat unexpected connection with schemes that are optimal from modulation-theoretic point of view.

Before proceeding to our results, we mention recent literature on the dispersion of wireless channels. Single antenna channel dispersion is computed in [7] for a coherent channel subject to stationary fading process. In [10] finite-blocklength effects are explored for the non-coherent block fading setup. Quasi-static fading channels in the general MIMO setting have been thoroughly investigated in [11], showing that the expansion (1) changes dramatically (in particular the channel dispersion term becomes zero); see also [12] for evaluation of the bounds. Coherent quasi-static channel has been studied in the limit of infinitely many antennas in [13] appealing to concentration properties of random matrices. Dispersion for lattices (infinite constellations) in fading channels has been investigated in a sequence of works, see [14] and references. Note also that there are some very fine differences between stationary and block-fading channel models, cf. [15, Section 4].

II. Main Results

This section is organized as follows: first, we characterize the set of capacity achieving input distributions for this channel, next we state the achievability theorem, then we compute $\mathrm{Var}[i(X;Y,H]|X]$ and give its minimizers over the set of capacity achieving input distributions.

A. Capacity achieving input distributions

It is instructive to first consider a special case of $n_t = T = 2$. As argued by Telatar [6], the following input achieves capacity

$$X = \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & \xi_3 \\ \xi_2 & \xi_4 \end{bmatrix}, \qquad (8)$$

where here and below ξ_j are i.i.d. standard normal random variables. Reflecting upon ingenious scheme of Alamouti [9] we observe that the following input is also capacity achieving:

$$X = \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & -\xi_2 \\ \xi_2 & \xi_1 \end{bmatrix} \qquad (9)$$

A computation shows that Alamouti achieves the smaller value of $\mathrm{Var}[i(X;Y,H)|X]$ and hence Telatar's i.i.d. Gaussian input (8) cannot be optimal from the dispersion point of view. But is there an input achieving yet a smaller variance than Alamouti?

As a first step towards the answer we characterize the set of all c.a.i.d.'s.

Proposition 1. *P_X is a capacity achieving input distribution iff X is jointly Gaussian and either of the following holds*

1) *Let R_i denote the i-th row of X, then:*

$$\mathbb{E}[R_i^T R_i] = \frac{P}{n_t} I_T, \qquad i = 1, \dots, n_t \qquad (10)$$

$$\mathbb{E}[R_i^T R_j] = -\mathbb{E}[R_j^T R_i], \qquad i \neq j \qquad (11)$$

2) *Let C_i be the i-th column of X, then:*

$$\mathbb{E}[C_i C_i^T] = \frac{P}{n_t} I_T, \qquad i = 1, \dots, T \qquad (12)$$

$$\mathbb{E}[C_i C_j^T] = -\mathbb{E}[C_j C_i^T], \qquad i \neq j \qquad (13)$$

Example: In the $n_t = T = 2$ case, the set of caids is given by

$$\left\{ \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & -\rho\xi_2 + \sqrt{1-\rho^2}\xi_3 \\ \xi_2 & \rho\xi_1 + \sqrt{1-\rho^2}\xi_4 \end{bmatrix} : -1 \le \rho \le 1 \right\} \qquad (14)$$

Where $\xi_1, \xi_2, \xi_3, \xi_4 \sim \mathcal{N}(0,1)$ i.i.d.

Remark 1. These conditions imply that if X is caid, then X^T and any submatrix of X are caids too (for different n_t and T). Elementwise, conditions require that all elements in a row are pairwise independent, all elements in a column are pairwise independent, each 2×2 minor has equal and opposite correlation across diagonal entries, and each of the entries have the same distribution $X_{ij} \sim \mathcal{N}(0, \frac{P}{n_t})$.

Proof Sketch: For this channel, P_X is a caid iff it induces the unique optimal output distribution $P^*_{Y|H}$. Using characteristic functions, we can show that in order to P_X to induce $P^*_{Y|H}$ for almost all h_0, our X must satisfy

$$h_0 X \sim \mathcal{N}\left(0, \frac{P}{n_t} \|h_0\|^2 I_T\right) \qquad (15)$$

The conditions stated above now follow from X

$$\mathbb{E}[(h_0 X)^T (h_0 X)] = \frac{P}{n_t} \left(\sum_{i=1}^{n_t} h_i^2 \right) \qquad (16)$$

■

B. Achievability bound

Here we give the coding theorem characterizing non-asymptotic achievable rates for this channel.

When X is a *full rate orthogonal design* (discussed below), we have a trivial proof of achievability: use the corresponding linear decoder to transform the channel into an equivalent SISO block fading channel, then apply the results from [7].

Theorem 2. *For the coherent MISO fading channel described, for any $n_t \times T$ caid X,*

$$\log M^*(nT, \epsilon, P) \ge nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + o(\sqrt{n}) \qquad (17)$$

Where $C(P)$ is the capacity (6) *and $V(P) = \frac{1}{T}\mathrm{Var}[i(X;Y,H)|X]$ is the conditional variance of the information density.*

Proof Sketch: Apply the $\kappa\beta$ bound from [1], choosing Q_Y to be the (unique) capacity achieving output distribution, and choosing the set F to be, for arbitrary $\delta > 0$,

$$F = \text{support}\left\{\sqrt{nTP}\frac{X^n}{\|X^n\|_F}\right\} \cap \{x^n : \mathrm{Var}(i(x^n;Y^n,H^n)) \leq V(P) + \delta\} \tag{18}$$

Where X is any caid. The proof proceeds similarly to [7]. ■

The natural question is now: what is the expression for $\frac{1}{T}\mathrm{Var}[i(X;Y,H)|X]$, and which caids X give its smallest value? These questions are answered next.

C. Information density and conditional variance

Note that all caid's P_X induce the same capacity achieving output distribution (c.a.o.d.) equal to $\mathcal{N}(0, I_T(1 + \frac{P\|H\|^2}{n_t}))$ (given H). Thus, the information density is the same for all caids P_X and is equal to

$$i(x;y,h) = \frac{T}{2}\log\left(1 + \frac{P\|h\|^2}{n_t}\right) + \frac{1}{2}\frac{\|hx\|^2 + 2hxZ^T - \frac{P}{n_t}\|h\|^2\|Z\|^2}{1 + \frac{P\|h\|^2}{n_t}}, \tag{19}$$

where $Z = hx - y$ is a $1 \times T$ real vector.

Using the information density, we can now compute the conditional variance for this channel.

Proposition 3. *For the MISO $n_t \times T$ block fading CSIR channel and capacity achieving input X the conditional variance is given by*

$$\frac{1}{T}\mathrm{Var}[i(X;Y,H)|X] = V_1(P) - \frac{\chi_2}{n_t^2 T}\mathrm{Var}(\|X\|_F^2) \tag{20}$$

where $V_1(P)$ is independent of the capacity achieving input distribution X, and

$$V_1(P) \triangleq T\mathrm{Var}\left(C_{AWGN}\left(\frac{P}{n_t}\|H\|^2\right)\right) + \mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\|H\|^2\right)\right] + 2\chi_1\left(\frac{P}{n_t}\right)^2 \tag{21}$$

$$\chi_1 \triangleq \mathbb{E}\left[\left(\lambda\left(\frac{P}{n_t}\|H\|^2\right)\|H\|^2\right)^2\right] \tag{22}$$

$$\chi_2 \triangleq \mathbb{E}^2\left[\lambda\left(\frac{P}{n_t}\|H\|^2\right)\|H\|^2\right] \tag{23}$$

$$V_{AWGN}(P) \triangleq \frac{1}{2}\left(1 - \left(\frac{1}{1+P}\right)^2\right) \tag{24}$$

$$\lambda(P) \triangleq \frac{1}{2(1+P)} = \frac{dC_{AWGN}(P)}{dP} \tag{25}$$

The last three quantities are the AWGN capacity, dispersion, and optimal Lagrange multiplier $\lambda(P)$ in the optimization

$$\sup_{P_X:\mathbb{E}[\|X\|^2]\leq P} I(X;Y) \tag{26}$$

Proof Sketch: Decompose variance using the identity

$$\mathbb{E}_X\mathrm{Var}[i(X;Y,H)|X] = \mathbb{E}_X\mathbb{E}_H\mathrm{Var}[i(X;Y,H)|X,H] + \mathbb{E}_X\mathrm{Var}_H\mathbb{E}[i(X;Y,H)|X,H] \tag{27}$$

The first term is simple to evaluate, and gives

$$T\mathbb{E}_H\left[V_{AWGN}\left(\frac{P}{n_t}\|H\|^2\right)\right] \tag{28}$$

The second term is more involved. The key identity is that the inner $\mathbb{E}[i(X;Y,H)|H,X]$ is equal to

$$TC_{AWGN}\left(\frac{P}{n_t}\|H\|^2\right) + \lambda\left(\frac{P}{n_t}\|H\|^2\right)\left(\|HX\|^2 - \frac{TP}{n_t}\|H\|^2\right) \tag{29}$$

Use this in the second term in (27) to get the remaining terms in (20). ■

Next we isolate the dependence of this quantity on the input distribution X, and analyze which input distributions are minimizers.

D. Capacity-dispersion optimal input distributions

In this section, we are interested in the minimizers to the conditional variance given in Proposition3.

Definition 1. For the MISO channel with n_t transmit antennas and coherence time T we define

$$v^*(n_t, T) \triangleq \max_{P_X:I(X;Y,H)=C} \sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{T}\sum_{l=1}^{T}\rho_{ikjl}^2 \tag{30}$$

where

$$\rho_{ikjl} = \frac{\mathrm{Cov}(X_{ik}, X_{jl})}{\sqrt{\mathrm{Var}(X_{ik})\mathrm{Var}(X_{jl})}} \tag{31}$$

Proposition 4. *The minimal dispersion of an $n_t \times T$ block-fading MISO channel is given by*

$$V_{min} \triangleq \inf_{X\text{-}caid} \frac{1}{T}\mathrm{Var}[i(X;Y,H)|X] = V_1(P) - \frac{2\chi_2 P^2}{n_t^4 T}v^*(n_t, T)$$

where V_1 and χ_2 are from (20)*.*

Proof: The only term that depends on X in (20) is $\mathrm{Var}[\|X\|_F^2]$. Since X is necessarily jointly Gaussian (Proposition 1), we have an easily verifiable identity:

$$\mathrm{Var}(\|X\|_F^2) = \sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{T}\sum_{l=1}^{T}\mathrm{Cov}(X_{ik}^2, X_{jl}^2) = 2\left(\frac{P}{n_t}\right)^2\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{T}\sum_{l=1}^{T}\rho_{ikjl}^2 \tag{32}$$

■

Therefore optimizing dispersion is equivalent to *maximizing* the amount of correlation amongst the entries of X. In a sense, this asks for the capacity achieving input distribution having the least amount of randomness.

We proceed to characterizing quantity $v^*(n_t, T)$. For convenience, we state a simple bound:

Lemma 5. *Let $R = (R_1, \dots, R_n) \sim \mathcal{N}(0, PI_n)$ and $A \sim N(0, P)$ be arbitrarily correlated with R, then*

$$\sum_{i=1}^{n} \rho^2_{AR_i} \le 1 \tag{33}$$

Where $\rho_{AR_i} = \mathrm{Cov}(A, R_i)/\mathrm{Var}(A) = \mathbb{E}[AR_i]/P$. We have equality iff $A = \rho_{AR_1} R_1 + \dots + \rho_{AR_n} R_n$.

Proof: Geometrically, projection of a unit vector A onto an orthonormal basis $\mathrm{Span}\{R_1, \dots, R_n\}$ has length at most 1, with equality iff $A \in \mathrm{Span}\{R_1, \dots, R_n\}$. ∎

Next we need to recall some facts about orthogonal designs. A real $n \times n$ orthogonal design of size k is defined to be an $n \times n$ matrix A with entries given by linear polynomials in $x_1, \dots, x_k$ and coefficients in $\mathbb{R}$ satisfying

$$A^T A = \left(\sum_{i=1}^{k} x_i^2 \right) I_n \tag{34}$$

Orthogonal designs may be represented as the sum $A = \sum_{i=1}^{k} x_i V_i$ where $\{V_1, \dots, V_k\}$ is a collection of $n \times n$ real matrices satisfying Hurwitz-Radon conditions:

$$V_i^T V_i = I_n\,, \quad V_i^T V_j + V_j^T V_i = 0 \quad i \neq j \tag{35}$$

The maximal size of an $n \times n$ real orthogonal design is called a Hurwitz-Radon number $\rho(n)$. From the classical work [16], [17] it is known:

$$\rho(2^a b) = 8 \left\lfloor \frac{a}{4} \right\rfloor + 2^{a \bmod 4}, \qquad a, b \in \mathbb{Z}, \text{b–odd}\,.$$

In particular, $\rho(n) \le n$ and $\rho(n) = n$ only for $n = 1, 2, 4, 8$.

Theorem 6. *For any pair of positive integers n_t, T we have*

$$v^*(T, n_t) = v^*(n_t, T) \le n_t T \min(n_t, T)\,. \tag{36}$$

Furthermore, the bound (36) *is tight if and only if $n_t \le \rho(T)$ or $T \le \rho(n_t)$.*

Proof: $v^*(n_t, T) = v^*(T, n_t)$ follows from the symmetry to transposition of caid-conditions on X (see Proposition 1) and symmetry to transposition of (30). From now on, without loss of generality we assume $n_t \le T$.

For the upper bound, we apply Lemma 5 to the rows (or columns) of X, giving

$$\sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{T} \sum_{l=1}^{T} \rho^2_{ijkl} \le \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{T} 1 = n_t T \min(n_t, T) \tag{37}$$

Note that (37) implies that if X achieves the bound (36), then removing the last row of X achieves (36) as an $(n_t - 1) \times T$ design. In other words, if (36) is tight for $n_t \times T$ then it is tight for all $n_t' \le n_t$.

With this observation in mind, take $n_t = \rho(T)$ and a maximal Hurwitz-Radon family $\{V_i, i = 1, \dots, n_t\}$ of $T \times T$ matrices. Let $\xi \sim \mathcal{N}(0, I_T)$ be i.i.d. normal row-vector. Consider

$$X = \left[V_1^T \xi^T \ \cdots V_{n_t}^T \xi^T \right]^T \tag{38}$$

The definition of orthogonal design (35) implies that rows of X satisfy conditions (10)-(11). Thus X is capacity achieving. On the other hand, it is easy to show in the representation (38) we have

$$\sum_{i,j,k,l} \rho^2_{ijkl} = \mathrm{tr}((\sum_{i=1}^{n_t} V_i V_i^T)^2) \tag{39}$$

Since V_i are orthogonal we have $V_i V_i^T = I_T$ and hence the trace above equals $n_t^2 T$, matching (36).

Conversely, suppose X is capacity achieving and attains (36). Again, we may represent X via (38) using $\xi \sim \mathcal{N}(0, I_d)$ where $d \ge T$. By conditions (10)-(11) we have that $d \times T$ matrices $\{V_i\}$ must satisfy equations (35). If $d = T$ then they constitute an orthogonal design and by Hurwitz-Radon we must have $n_t \le \rho(T)$. If $d > T$ then there must exist $X_{i,j}$ which is not a linear combination of $\{X_{1,s}, s = 1, \dots, T\}$ (i.e. not in the span), thus X cannot attain the bound (37), contradicting the assumption. ∎

E. Discussion

Elementary results on orthogonal designs show that conditions for tightness of (36) are satisfied if and only if a full rate real orthogonal design of dimensions $n_t \times T$ or $T \times n_t$ exists, cf. [8] or [18, Proposition 4]. Consequently, each full-rate orthogonal design yields a caid X that achieves minimal dispersion. Some examples (ξ_j are i.i.d. $\mathcal{N}(0,1)$)

$$X = \sqrt{\frac{P}{4}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \xi_4 \\ -\xi_2 & \xi_1 & -\xi_4 & \xi_3 \\ -\xi_3 & \xi_4 & \xi_1 & -\xi_2 \\ -\xi_4 & -\xi_3 & \xi_2 & \xi_1 \end{bmatrix} \tag{40}$$

$$X = \sqrt{\frac{P}{4}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & -\xi_4 \\ -\xi_3 & \xi_4 & \xi_1 \\ -\xi_4 & -\xi_3 & \xi_2 \end{bmatrix} \tag{41}$$

Note that for the designs with entries $\pm\xi_j$, with ξ_j being independent Gaussians, computation of the sum (30) is simplified:

$$\sum_{ijkl} \rho^2_{ikjl} = \sum_{t=1}^{d} (\ell_t)^2\,, \tag{42}$$

where ℓ_t is the number of times $\pm\xi_t$ appears in the description of X. By this observation and the remark after Proposition 1 we can obtain lower bounds on $v^*(n_t, T)$ for $n_t > \rho(T)$ via the following *truncation construction*:

1) Take $T' > T$ such that $\rho(T') \ge n_t$ and let X' be a corresponding $\rho(T') \times T'$ full-rate orthogonal design (with entries $\pm\xi_1, \dots \pm \xi_{T'}$).
2) Choose an $n_t \times T$ submatrix of X' maximizing the sum of squares of the number of occurrences of each of ξ_j, cf. (42).

As an example of this method, by truncating a 4×4 design (41) we obtain the following 2×3 and 3×3 submatrices:

$$X = \sqrt{\frac{P}{3}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & \xi_4 \\ -\xi_3 & -\xi_4 & \xi_1 \end{bmatrix} \quad X = \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & \xi_4 \end{bmatrix} \tag{43}$$

TABLE I
VALUES FOR $v^*(n_t, T)$

$n_t \setminus T$	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	8
2		8	10	16	18	24	26	32
3			21	36	[39,45)	[46,54)	[57,63)	72
4				64	[68,80)	[80,96)	[100,112)	128
5					[89,125)	[118,150)	[155,175)	200
6						[168,216)	[222,252)	288
7							[301,343)	392
8								512

Note: Table is symmetric about diagonal; intervals $[a, b)$ mark entries for which dispersion-optimal input is unknown.

TABLE II
VALUES OF $\frac{V_{min}}{C^2}$ (WHEN KNOWN) AT $SNR = 20\ dB$

$n_t \setminus T$	1	2	3	4	5	6	7	8
1	0.38	0.60	0.82	1.05	1.27	1.49	1.72	1.94
2	0.35	0.39	0.52	0.60	0.73	0.82	0.94	1.03
3	0.38	0.42	0.45	0.49				0.79
4	0.42	0.43	0.44	0.45				0.69
5	0.46	0.48						0.64
6	0.50	0.51						0.62
7	0.54	0.55						0.61
8	0.59	0.59	0.59	0.60	0.60	0.60	0.61	0.61

By independent methods we were able to show that designs (43) are dispersion-optimal and attain $v^*(3, 3) = 21$ and $v^*(2, 3) = 10$. Note that in these cases the bound (36) is not tight, illustrating the "only if" part of Theorem 6.

Orthogonal designs were introduced into communication theory by Tarokh et al [8] as a natural generalization of Alamouti's scheme [9]. In cases when full-rate designs do not exist, there have been various suggestions as to what could be the best solution, e.g. [18]. Thus for non full-rate designs the property of minimizing dispersion (such as (43)) could be used for selecting the best design for cases $n_t > \rho(T)$.

Our current knowledge about v^* is summarized in Table I. The lower bounds for cases not handled by Theorem 6 were computed by truncating the 8x8 orthogonal design [8, (5)]. Based on the evidence from $2 \times T$ and 3×3 we *conjecture* this construction to be optimal.

Finally, returning to the original question of the minimal delay required to achieve capacity, see (2), we calculate the value of $\frac{V_{min}}{C^2}$ in Table II.

From the proof of Theorem 6 it is clear that Telatar's i.i.d. Gaussian (as in (8)) is never dispersion optimal, unless $n_t = 1$ or $T = 1$. Indeed, for Telatar's input $\rho_{ikjl} = 0$ unless $(i, k) = (j, l)$. Thus embedding even a single Alamouti block (9) into an otherwise i.i.d. $n_t \times T$ matrix X strictly improves the sum (30).

We note that the value of $\frac{V}{C^2}$ entering (2) can be quite sensitive to the suboptimal choice of the design. For example, for $n_t = T = 8$ and $SNR = 20\ dB$ estimate (2) shows that one needs

- around 600 channel inputs (that is 600/8 blocks) for the optimal 8×8 orthogonal design, or
- around 850 channel inputs for Telatar's i.i.d. Gaussian design

in order to achieve 90% of capacity. This translates into a 40% longer delay (or battery spent in running the decoder) with unoptimized transmitter.

Thus, curiously even in cases where pure multiplexing (that is maximizing transmission rate) is needed – as is often the case in modern cellular networks – transmit diversity enters the picture by enhancing the finite blocklength fundamental limits. We remind, however, that our discussion pertains only to cases when the transmitter (base-station) is equipped with more antennas than the receiver (user equipment).

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," *Eur. trans. telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

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

[8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[9] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[10] 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.

[11] ——, "Quasi-static MIMO fading channels at finite blocklength," *IEEE Trans. Inform. Theory*, To appear, 2014.

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

[13] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO Rayleigh block-fading channel," *IEEE Trans. Inf. Theory*, 2013, submitted. [Online]. Available: http://arxiv.org/abs/1303.3400

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

[15] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 2003 41st Allerton Conference*, vol. 41, no. 1, Allerton Retreat Center, Monticello, IL, USA, 2003, pp. 410–419.

[16] A. Hurwitz, "Über die komposition der quadratischen formen," *Math. Ann.*, vol. 88, no. 1, pp. 1–25, 1922.

[17] J. Radon, "Lineare scharen orthogonaler matrizen," in *Abh. Sem. Hamburg*, vol. 1. Springer, 1922, pp. 1–14.

[18] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2468–2503, Oct. 2003.