RECURSIVE TRAINING WITH UNITARY MODULATION FOR CORRELATED BLOCK-FADING MIMO CHANNELS

Wenyi Zhang, Student Member, IEEE, and J. Nicholas Laneman, Member, IEEE

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN, USA

ABSTRACT

We propose and analyze a communication scheme called recursive training with unitary modulation (RTUM) for correlated block fading channels with multiple inputs and multiple outputs (MIMO) and without channel state information (CSI). In RTUM, we employ unitary matrices as constellation symbols, and utilize a recursive procedure to perform channel estimation and demodulation/decoding in an alternating manner. Central benefits of RTUM include: low-complexity implementation; transformation of the original fading channel without CSI into a series of parallel sub-channels, each with perfect receive CSI but additional Gaussian noise; and convenience of analysis. By inspecting the effective signal-to-noise ratio (SNR) of the RTUM, we are further able to identify three operating regimes that lead to distinct channel behaviors. For typical channels we establish a new rule of thumb that spectrally-efficient low-SNR "coherent" communication can be realized when the average channel SNR is larger than the mean square one-step prediction error of the block fading process.

1. INTRODUCTION

For time-varying fading channels, explicit channel estimation using pilot symbols has long been a popular practice; see [1] and references therein. Compared to approaches that adopt non-coherent techniques [2], pilot-assisted schemes are pseudo-coherent, and often simpler to implement. In a typical pilot-assisted scheme, a sequence of pilot symbols known to the receiver is inserted sparingly among data symbols. Based upon the received pilot symbols, the receiver forms an estimate of the fading process, which is then treated as the true fading process to facilitate coherent reception.

Since channel estimation cannot be error-free, channel inputs affect the effective noise of the trained channel, generally making it non-Gaussian and input-dependent. This imposes a fundamental difficulty in the analysis of pilot-assisted schemes. Considerable effort has been devoted to this issue; see [3]-[6] and references therein. This previous research mainly focuses on the regime of high signal-to-noise ratio (SNR) under circular complex Gaussian inputs that are known to be capacity-achieving for channels with perfect channel state information (CSI). A typical rule of thumb is that communication can be viewed as essentially "coherent", *i.e.*, having perfect receive CSI, if the average channel SNR is less than the reciprocal of the mean square one-step prediction error of the fading process.

In this paper we seek to resolve the above difficulty from another route. To this end, we propose and analyze a scheme called recursive training with unitary modulation (RTUM). The basic idea is that, by choosing channel inputs to be unitary matrices instead of circular complex Gaussian, the resulting effective noise of the trained channel becomes both circular complex Gaussian and independent of the channel inputs. Furthermore, via a recursive procedure that performs channel estimation and demodulation/decoding in an alternating manner, RTUM converts the original fading channel without CSI into a series of parallel sub-channels each with perfect receive CSI and with additional Gaussian noise.

The recursive training structure in RTUM is similar to that used in [6]; in fact, from a high-level perspective it can be viewed as an instance of the decision-feedback decoder for general finite-state Markov channels [7]. In RTUM, a key point is that unitary constellations lead to a recursive procedure that can be implemented with very low complexity, and the additional noise due to channel estimation remains Gaussian. In the channel estimation phase, only the classical Kalman filter [8] is needed, and, because the filter coefficients do not involve the actual transmitted data, they can be computed off-line. In the demodulation/decoding phase, the RTUM-trained channel is precisely linear Gaussian, hence standard maximum-likelihood (ML) decoding is readily applicable, rather than mismatched decoding [3]. Furthermore, in RTUM as the interleaving depth grows the decomposed parallel sub-channels monotonically converge

This work has been supported in part by the State of Indiana through the 21st Century Research and Technology Fund, and by the National Science Foundation through contract ECS03-29766.

to a limiting channel, hence leading to a further simplification for performance analysis.

Analyzing the limiting channel, we identify three operating SNR regimes that lead to qualitatively distinct channel behaviors. The regimes provide a unified perspective for time-varying block fading channels. For typical slowly time-varying channels, we observe that RTUM is able to convert the original fading channel without CSI into a fading channel with perfect receive CSI and essentially the same SNR, as long as the channel SNR is larger than the mean square one-step prediction error ϵ of the block fading process, and is smaller than its reciprocal $1/\epsilon$. The later threshold $1/\epsilon$ corresponds to the rule of thumb from previous work mentioned above, and the low SNR threshold ϵ is new.

The remainder of the paper is organized as follows. Section 2 summarizes the fading channel model. Section 3 describes the RTUM communication scheme and derives the trained channel. Section 4 presents a performance analysis, with a focus on the general effective SNR of the trained channel as well as achievable rates at low SNR. Finally, Section 5 concludes the paper. An Appendix includes some details of the necessary derivations. All logarithms are taken with base e, and information is measured in units of nats.

2. CHANNEL MODEL

We consider a system with M transmit and N receive antennas. Transmit-receive antenna pairs suffer spatially independent, identically distributed (i.i.d.) frequency non-selective Rayleigh fading, and receive antennas suffer temporally and spatially i.i.d. circular complex Gaussian additive noise. We adopt a block Gauss-Markov channel (BGMC) model. In a BGMC, the realizations of the fading coefficients remain constant in a block of channel uses, and change to new but correlated realizations in the next block. For the sake of analysis, we further set the block length to be equal to M, the number of transmit antennas. Hence we can write the channel equation in length-M blocks as

$$\mathbf{X}(l) = \sqrt{\rho} \cdot \mathbf{S}(l) \cdot \mathbf{H}(l) + \mathbf{Z}(l), \quad (1)$$

where the index l denotes the lth block of channel uses, and ρ denotes the average channel SNR per channel use. $\mathbf{S} \in \mathcal{C}^{M \times M}$ represents the matrix channel input, normalized such that

$$\frac{1}{M} \cdot \mathcal{E}\{\mathrm{tr}\{\mathbf{S}\mathbf{S}^{\dagger}\}\} = 1.$$

$\mathbf{X} \in \mathcal{C}^{M \times N}$ represents the corresponding channel output. The additive noise matrix $\mathbf{Z} \in \mathcal{C}^{M \times N}$ has i.i.d. $\mathcal{CN}(0,1)$ elements, and is also independent for different l. The fading coefficient matrix $\mathbf{H} \in \mathcal{C}^{M \times N}$ has spatially i.i.d. $\mathcal{CN}(0,1)$ elements, and is temporally correlated according to a first-order Gauss-Markov state evolution equation as

$$\mathbf{H}(l+1) = \sqrt{1-\epsilon} \cdot \mathbf{H}(l) + \sqrt{\epsilon} \cdot \mathbf{V}(l+1). \quad (2)$$

To ensure stationarity and ergodicity, the innovation rate ϵ satisfies $0 < \epsilon \leq 1$; $\mathbf{V} \in \mathcal{C}^{M \times N}$ has i.i.d. $\mathcal{CN}(0,1)$ elements and is independent for different l; and the initial fading coefficient matrix $\mathbf{H}(0)$ also has i.i.d. $\mathcal{CN}(0,1)$ elements and is independent of $\mathbf{V}(l)$ for all l.

As described by (2), the BGMC is a special case of the channel model adopted in [9], mixing the generic block fading model [2] and the generic Gauss-Markov fading model [6]. Specifically, for the single transmit antenna case ($M = 1$), the BGMC reduces to the generic Gauss-Markov fading model. Admittedly, the BGMC is an artificial and approximate model in nature. By letting $\mathbf{H}$ remain constant for a block of M channel uses, we ignore the channel fluctuation within the block. However, practical values of M and ϵ suggest that the approximation may be reasonable. To see this, consider a channel described by a generic Gauss-Markov fading model with innovation rate ϵ'. Approximating it with a BGMC yields the innovation rate ϵ in (2) to be $\epsilon = 1 - (1-\epsilon')^M \approx M\epsilon'$ if $M\epsilon' \ll 1$. In practical systems the number of transmit antennas is usually small, *e.g.*, $M < 10$; furthermore, the fading process is typically underspread, with ϵ ranging from $1.8 \cdot 10^{-2}$ to $3 \cdot 10^{-7}$ [6].

We note that care should be taken in extending the BGMC to certain extreme cases, in which the ignored intra-block channel fluctuation may become significant. For instance it does not make sense to let M grow to infinity. Furthermore, we will avoid asymptotic analysis as $\rho \to \infty$, since in that regime a fundamental discrepancy arises between BGMC (for $M > 1$) whose capacity grows logarithmically [10] [9], and the generic Gauss-Markov model whose capacity grows double-logarithmically [11].

3. RECURSIVE TRAINING WITH UNITARY MODULATION (RTUM)

Figure 1 illustrates the interleaving scheme used by RTUM. Each symbol corresponds to a block of M channel uses, as described by (1). Similar to that used in [6], the interleaving scheme decomposes the channel into L parallel sub-channels (PSC). The lth ($l = 0, 1, \ldots, L-1$) PSC includes all the $(k \cdot L + l)$th blocks of the original BGMC for $k = 0, 1, \ldots, K-1$. The L PSCs suffer correlated fading, and this correlation is exactly what we seek to exploit using RTUM. Although some residual correlation remains within each PSC among its K block symbols, for an ergodic BGMC, this correlation vanishes as the interleaving depth $L \to \infty$. In practical systems with finite L, if necessary, we can utilize an additional interleaver for each PSC to make it essentially memoryless.

We make a slight change of notations in the sequel. Since all the PSCs are memoryless, when describing a PSC we can simply suppress the internal index k among its K coding block symbols, and only indicate the PSC index l without

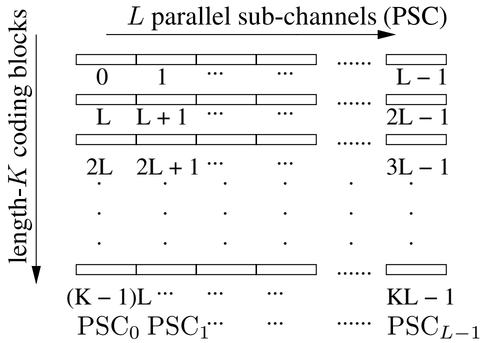

Fig. 1. The interleaving scheme used by RTUM.

loss of generality. For example, $\mathbf{H}(l)$ actually corresponds to any of $\mathbf{H}(k \cdot L + l)$ for $k = 0, 1, \ldots, K-1$.

In RTUM, each block symbol in (1) is an $M \times M$ unitary matrix, *i.e.*,

$$\mathbf{S}(l)\mathbf{S}^\dagger(l) = I_{M\times M}.$$

We note that for the case of a single transmit antenna, unitary constellations reduce to phase-shift keying (PSK).

In RTUM we utilize a block decision-feedback structure to exploit channel correlation, leading to a recursive implementation that performs channel estimation and demodulation/decoding in an alternating manner. To initialize transmission, PSC 0, the first parallel sub-channel, transmits a pilot sequence known to the receiver. Based upon the received pilot sequence in PSC 0, the receiver predicts $\mathbf{H}(1)$, the fading coefficient matrix of PSC 1, and proceeds to demodulate and decode the transmitted sequence in PSC 1 coherently. If the rate of PSC 1 does not exceed the corresponding channel mutual information, then information theory ensures that, for sufficient many channel uses, *i.e.*, as $K \to \infty$, there always exist codes that have arbitrarily small decoding error probability. Hence the receiver can, at least in principle, form an error-free reconstruction of the transmitted sequence in PSC 1, which then effectively becomes a "fresh" training sequence to facilitate the prediction of $\mathbf{H}(2)$ and subsequent coherent demodulation/decoding of PSC 2. Alternating the estimation-demodulation/decoding procedure repeatedly, all the PSCs are reliably decoded one after another.

By induction, let us consider PSC l, assuming that the block symbols $\{\mathbf{S}(i)\}_{i=0}^{l-1}$ of the previous PSCs have all been successfully reconstructed at the receiver. These previous PSCs can then be written as

$$\begin{aligned} \mathbf{H}(i+1) &= \sqrt{1-\epsilon}\cdot\mathbf{H}(i) + \sqrt{\epsilon}\cdot\mathbf{V}(i+1) \quad (3)\\ \mathbf{X}(i) &= \sqrt{\rho}\cdot\mathbf{S}(i)\cdot\mathbf{H}(i) + \mathbf{Z}(i), \quad (4)\end{aligned}$$

for $i = 0, 1, \ldots, l-1$. Since $\{\mathbf{S}(i)\}_{i=0}^{l-1}$ are unitary matrices known to the receiver, the observation equation (4) can be further rewritten as

$$\begin{aligned} \mathbf{S}^\dagger(i)\cdot\mathbf{X}(i) &= \sqrt{\rho}\cdot(\mathbf{S}^\dagger(i)\mathbf{S}(i))\cdot\mathbf{H}(i) + \mathbf{S}^\dagger(i)\cdot\mathbf{Z}(i)\\ &= \sqrt{\rho}\cdot\mathbf{H}(i) + \mathbf{Z}'(i),\end{aligned}$$

where the rotated noise matrix $\mathbf{Z}'(i) \overset{\text{def}}{=} \mathbf{S}^\dagger(i)\mathbf{Z}(i)$ also has i.i.d. $\mathcal{CN}(0,1)$ elements. Utilizing the Kalman filter [8], the one-step minimum mean-square error (MMSE) prediction of $\mathbf{H}(l)$ is recursively given by

$$\begin{aligned} \hat{\mathbf{H}}(l) &\overset{\text{def}}{=} \mathcal{E}\left\{\mathbf{H}(l) \mid \{\mathbf{S}^\dagger(i)\mathbf{X}(i)\}_{i=0}^{l-1}\right\}\\ &= \left(\sqrt{1-\epsilon} - \sqrt{\rho}\cdot K(l-1)\right)\cdot\hat{\mathbf{H}}(l-1)\\ &\quad + K(l-1)\cdot\mathbf{S}^\dagger(l-1)\mathbf{X}(l-1), \quad (5)\end{aligned}$$

where the scalar gain $K(l-1)$ and the prediction error variance $\sigma^2(l)$ are

$$\begin{aligned} K(l-1) &= \frac{\sigma^2(l-1)\cdot\sqrt{(1-\epsilon)\cdot\rho}}{\sigma^2(l-1)\cdot\rho+1} \quad (6)\\ \sigma^2(l) &= \frac{(1-\epsilon)\cdot\sigma^2(l-1)}{\sigma^2(l-1)\cdot\rho+1} + \epsilon. \quad (7)\end{aligned}$$

The recursions (6) and (7) are initialized by $\sigma^2(0) = 1$ and $\hat{\mathbf{H}}(0) = 0_{M\times N}$. The prediction error variance $\sigma^2(l)$ is defined for each individual element of $\mathbf{H}(l)$, and is the same for all these elements, since the fading process is modeled as spatially i.i.d.

For MMSE estimation, a convenient property is that the estimate $\hat{\mathbf{H}}(l)$ and the estimation error $\tilde{\mathbf{H}}(l) \overset{\text{def}}{=} \mathbf{H}(l) - \hat{\mathbf{H}}(l)$ matrices have i.i.d. $\mathcal{CN}\left(0, 1-\sigma^2(l)\right)$ and $\mathcal{CN}\left(0, \sigma^2(l)\right)$ elements, respectively, and they are uncorrelated thus independent. The channel equation of PSC l then becomes

$$\begin{aligned} \mathbf{X}(l) &= \sqrt{\rho}\cdot\mathbf{S}(l)\cdot\mathbf{H}(l) + \mathbf{Z}(l)\\ &= \sqrt{\rho}\cdot\mathbf{S}(l)\cdot\hat{\mathbf{H}}(l) + \underbrace{\sqrt{\rho}\cdot\mathbf{S}(l)\cdot\tilde{\mathbf{H}}(l) + \mathbf{Z}(l)}_{\bar{\mathbf{Z}}(l)} \quad (8)\end{aligned}$$

where the effective noise matrix $\bar{\mathbf{Z}}(l)$ has i.i.d. circular complex Gaussian elements, and is independent of both the channel input $\mathbf{S}(l)$ and the estimated fading coefficient matrix $\hat{\mathbf{H}}(l)$. The effective SNR of channel (8) is

$$\rho(l) = \frac{1-\sigma^2(l)}{\sigma^2(l)\cdot\rho+1}\cdot\rho. \quad (9)$$

4. PERFORMANCE ANALYSIS

For each PSC with perfect receive CSI as described by (8), its mutual information $\mathcal{I}\left(\mathbf{S}(l);\mathbf{X}(l)|\hat{\mathbf{H}}(l)\right)$ is determined once the operating SNR (9) and the distribution of the input $\mathbf{S}(l)$ are given. In total, the achievable rate of RTUM is

$$R = \frac{1}{ML}\sum_{l=0}^{L}\mathcal{I}\left(\mathbf{S}(l);\mathbf{X}(l)|\hat{\mathbf{H}}(l)\right). \quad (10)$$

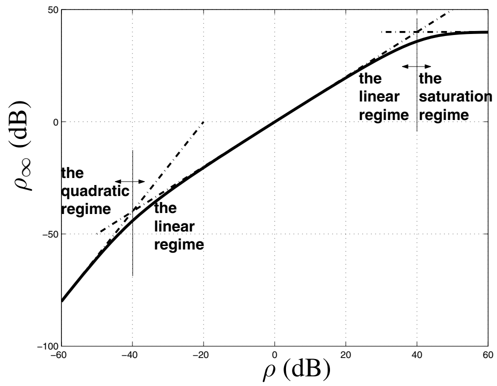

Fig. 2. Illustration of the three operating regimes, $\epsilon = 10^{-4}$.

We can evaluate and optimize (10) numerically, say, using simple Monte Carlo methods.

In the following we focus on the limiting behavior of R as $L \to \infty$. For all $\rho > 0$ and $0 < \epsilon \leq 1$, the sequence $\{\sigma^2(l)\}_{l=0}^{\infty}$ in (7) monotonically converges to [8]

$$\sigma_\infty^2 \stackrel{\text{def}}{=} \lim_{l\to\infty} \sigma^2(l) = \frac{(\rho-1)\cdot\epsilon + \sqrt{(\rho-1)^2\cdot\epsilon^2 + 4\cdot\rho\cdot\epsilon}}{2\cdot\rho}.$$

Consequently the sequence $\{\rho(l)\}_{l=0}^{\infty}$ in (9) converges to

$$\rho_\infty \stackrel{\text{def}}{=} \lim_{l\to\infty} \rho(l) = \frac{1-\sigma_\infty^2}{\sigma_\infty^2\cdot\rho+1}\cdot\rho. \tag{11}$$

Based upon (11), for small ϵ we identify three qualitatively distinct operating regimes as follows.

- *The quadratic regime:*
  For $\rho \ll \epsilon$, $\quad \sigma_\infty^2 \approx 1 - \rho/\epsilon$, $\quad \rho_\infty \approx \rho^2/\epsilon$;
- *The linear regime:*
  For $\epsilon \ll \rho \ll 1/\epsilon$, $\quad \sigma_\infty^2 \approx \sqrt{\epsilon/\rho}$, $\quad \rho_\infty \approx \rho$;
- *The saturation regime:*
  For $1/\epsilon \ll \rho$, $\quad \sigma_\infty^2 \approx \epsilon$, $\quad \rho_\infty \approx 1/\epsilon$.

We illustrate the three regimes in Figure 2, for $\epsilon = 10^{-4}$. The different slopes of ρ_∞ for the three regimes are clearly indicated. The linear regime covers roughly 80 dB, from -40 dB to $+40$ dB.

Both the quadratic regime and the saturation regime are undesirable for RTUM. In the quadratic regime, the quality of channel estimation is low, *i.e.*, $\sigma_\infty^2 \approx 1$. Instead, on-off signaling with unbounded peak power is asymptotically optimal in the zero-SNR limit [13]. In the saturation regime, channel estimation meets its fundamental limit ϵ, which cannot be further decreased by increasing SNR. Instead, variable signal amplitudes become the key resulting in unbounded mutual information as SNR grows [11].

An interesting observation is that, the two thresholds dividing the three regimes are determined by a single parameter ϵ, which can be viewed as the fundamental limit on the variance of the one-step prediction error of the fading process. The $1/\epsilon$ threshold dividing the linear and the saturation regimes coincides with that obtained in [6], in which it is derived for a generic Gauss-Markov fading model. In [6] it is shown that for the regime of $0 \ll \rho \ll 1/\epsilon$, circular complex Gaussian inputs with nearest-neighbor decoding achieve the full multiplexing gain $\min\{M, N\}$. By contrast, RTUM results in a penalty in the multiplexing gain at high SNR, due to the limitation of unitary constellations. For a scalar linear Gaussian channel without fading, PSK leads to an achievable rate $\frac{1}{2}\log\rho + O(1)$ as $\rho \to \infty$ [12]. To our knowledge, for coherent multi-antenna fading channels no asymptotic analysis is known for unitary constellations.

For time-varying fading channels without CSI under peak-power limited inputs like unitary constellations, previous asymptotic analysis [13] shows that the channel mutual information vanishes according to $O(\rho^2)$ as $\rho \to 0$. We revisit this behavior in the quadratic regime. Furthermore, the $\epsilon/(1-\epsilon)$ threshold dividing the quadratic and the linear regimes clearly indicates when this low-SNR asymptotic channel behavior becomes dominant. For typical ϵ, above $\rho \approx \epsilon$, we essentially have a low-SNR fading channel with perfect receive CSI. Applying the second-order mutual information expansion formula [14], we show in the Appendix that the total achievable rate R (10) approaches the following limit as $L \to \infty$,

$$\lim_{L\to\infty} R = \underbrace{\frac{\rho_\infty}{\rho}\cdot\left(1-\frac{M+N}{2M}\cdot\rho_\infty\right)}_{\mathcal{D}}\cdot(N\cdot\rho) + o(\rho^2) \tag{12}$$

under mild additional constraints on channel input $\mathbf{S}$.

Ignoring the higher-order term $o(\rho^2)$ and noticing that $(N\cdot\rho)$ corresponds to the low-SNR capacity limit as SNR $\rho \to 0$, the factor $\mathcal{D}$ then quantifies how closely the low-SNR capacity limit is approached. In Figure 3, we plot $\mathcal{D}$ vs. ρ/ϵ for different ϵ. Although all the curves vanish rapidly below the threshold $\rho/\epsilon = 1$, for $\rho/\epsilon > 1$, $\mathcal{D}$ can be reasonably close to 1. For instance, taking $\epsilon = 10^{-4}$, the optimal operating SNR is $\rho \approx 300\epsilon \approx -15$ dB, and the corresponding maximized $\mathcal{D}$ is above 0.9, *i.e.*, more than 90% of the low-SNR capacity limit is achievable.

5. CONCLUSION

This paper investigates a recursive pilot-assisted communication scheme (RTUM) to exploit the inherent correlation in time-varying fading channels without CSI. RTUM converts the original fading channel without CSI into a series of parallel sub-channels, each with perfect receive CSI. Coherent communication can be implemented with low complex-

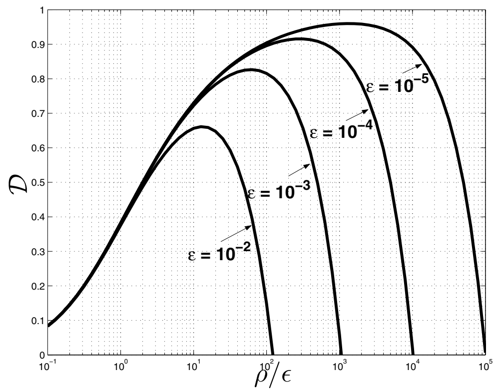

Fig. 3. The factor $\mathcal{D}$ vs. ρ/ϵ, $M = N$.

ity. Furthermore, performance analysis becomes essentially straightforward. We identify three operating regimes that lead to qualitatively distinct channel behaviors, providing a unified perspective for time-varying fading channels. The feasibility of spectrally-efficient low-SNR communication is established, complementing the previous related research pertaining the feasibility of essentially coherent communication at high SNR.

We note that RTUM is readily applicable to more general circular complex Gaussian fading processes. However, many issues remain open for further research, for example: the RTUM-trained channel behavior in the high SNR part of the linear regime; the effect of error propagation in the recursive receiver; the tradeoff between rate and delay for finite L; and efficient implementation of modulation/demodulation for unitary constellations.

6. APPENDIX: A SKETCH OF THE PROOF OF (12)

According to (10) and (11), as $L \to \infty$, $\lim_{L\to\infty} R$ converges to the mutual information of channel (8) operating at SNR ρ_∞, further divided by M, as we measure rates per channel use. We can rewrite (8) in the $L \to \infty$ limit as

$$\mathbf{X} = \sqrt{\rho_\infty} \cdot \mathbf{S} \cdot \mathbf{H} + \mathbf{Z},$$

with $\mathbf{H}$ and $\mathbf{Z}$ both normalized to have spatially i.i.d. $\mathcal{CN}(0,1)$ elements, and all the time indexes suppressed.

Utilizing the vectorization operator to convert the matrix channel into an equivalent vector channel, then applying [14, Theorem 3], we have

$$\lim_{L\to\infty} R = \frac{1}{M} \cdot \mathrm{tr}\left\{\mathcal{E}\left\{\Delta\right\}\right\} \cdot \rho_\infty - \frac{1}{2M} \cdot \mathrm{tr}\left\{\mathcal{E}\left\{\Delta^2\right\}\right\} \cdot \rho_\infty^2 + o(\rho_\infty^2),$$

for proper complex input $\mathbf{S}$, *i.e.*,

$$\mathcal{E}\left\{\mathrm{vec}\{\mathbf{S}\}\mathrm{vec}\{\mathbf{S}\}^{\mathrm{T}}\right\} = \mathcal{E}\left\{\mathrm{vec}\{\mathbf{S}\}\right\} \cdot \mathcal{E}\left\{\mathrm{vec}\{\mathbf{S}\}\right\}^{\mathrm{T}},$$

where

$$\Delta = (\mathbf{H}^{\mathrm{T}} \otimes I_{M\times M}) \cdot \mathrm{cov}(\mathrm{vec}\{\mathbf{S}\}) \cdot (\mathbf{H}^{\mathrm{T}} \otimes I_{M\times M})^{\dagger}.$$

If we further let the unitary input $\mathbf{S}$ have a scaled identity covariance matrix

$$\mathrm{cov}(\mathrm{vec}\{\mathbf{S}\}) = \frac{1}{M} \cdot I_{M^2\times M^2},$$

then Δ reduces to

$$\Delta = \frac{1}{M}(\mathbf{H}^{\dagger}\mathbf{H})^{\mathrm{T}} \otimes I_{M\times M}.$$

Noticing that we have normalized $\mathbf{H}$ to have spatially i.i.d. $\mathcal{CN}(0,1)$ elements, we can further show that

$$\begin{aligned} \mathrm{tr}\left\{\mathcal{E}\{\Delta\}\right\} &= \mathrm{tr}\left\{\mathcal{E}\{\mathbf{H}^{\dagger}\mathbf{H}\}\right\} = MN \\ \mathrm{tr}\left\{\mathcal{E}\{\Delta^2\}\right\} &= \frac{1}{M} \cdot \mathrm{tr}\left\{\mathcal{E}\{(\mathbf{H}^{\dagger}\mathbf{H})^2\}\right\} = N(M+N), \end{aligned}$$

and (12) consequently follows.

In the above derivations we have assumed that there exist distributions for the unitary input $\mathbf{S}$ such that $\mathrm{vec}\{\mathbf{S}\}$ is proper complex, and has a scaled identity covariance matrix. For small M, examples of such constellations can be explicitly constructed. In general, choosing $\mathbf{S}$ to be $M \times M$ isotropically distributed (i.d.) unitary matrices suffices to meet these constraints. For the definition and properties of i.d. unitary matrices we refer to [2] [10]. Intuitively, choosing $\mathbf{S}$ to be $M \times M$ i.d. unitary matrices corresponds to choosing $\mathbf{S}$ equally likely, or, uniformly, among all $M \times M$ unitary matrices. The constraints on $\mathbf{S}$ can then be verified by utilizing the symmetry in i.d. unitary matrices.

7. REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-Assisted Wireless Transmissions," *IEEE Mag. Signal Processing*, pp. 12–25, Nov. 2004.

[2] T. L. Marzetta and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[3] A. Lapidoth and S. Shamai (Shitz), "Fading Channels: How Perfect Need "Perfect Side Information" Be?," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[4] J. Baltersee, G. Fock, and H. Meyr, "Achievable Rate of MIMO Channels with Data-Aided Channel Estimation and Perfect Interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[5] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[6] R. Etkin and D. N. C. Tse, "Degrees of Freedom in Underspread MIMO Fading Channels," *IEEE Trans. Inform. Theory*, May 2003, submitted for publication.

[7] A. J. Goldsmith and P. P. Varaiya, "Capacity, Mutual Information, and Coding for Finite-State Markov Channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[8] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

[9] Y. Liang and V. V. Veeravalli, "Capacity of Noncoherent Time-Selective Rayleigh-Fading Channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[10] L. Zheng and D. N. C. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[11] A. Lapidoth and S. M. Moser, "Capacity Bounds via Duality with Applications to Multiple-Antenna Systems on Flat Fading Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2567, Oct. 2003.

[12] A. D. Wyner, "Bounds on Communication with Polyphase Coding," *Bell Sys. Tech. J.*, vol. 45, pp. 523–559, Apr. 1966.

[13] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[14] V. V. Prelov and S. Verdu, "Second-Order Asymptotics of Mutual Information," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.