Achieving Coherent Capacity of Correlated MIMO Channels in the Low-Power Regime with Non-Flashy Signaling Schemes

Vasanthan Raghavan and Akbar M. Sayeed
University of Wisconsin-Madison, Madison, WI, USA
raghavan@cae.wisc.edu, akbar@engr.wisc.edu

***Abstract*— Existing works on MIMO channel capacity assume that the channel power scales quadratically with antenna dimensions. In a recent work, we introduced a framework for sub-quadratic channel power scaling to model physical MIMO channels, and studied a family of channels with different spatial distributions of channel power. In particular, we introduced the notion of a beamforming channel, realizable in practice via the use of reconfigurable antenna arrays, that yields the highest capacity and maximal spectral efficiency in the low SNR regime. In this paper, we show that the MIMO beamforming channel requires a significantly smaller coherence requirement to achieve the wideband capacity limit at finite bandwidth than has been known in the literature. Contrasting the recent claims that training-based signaling schemes are *strictly* sub-optimal over communication schemes that do not perform explicit training, we show that training offers significant advantages in achieving wideband capacity. The price of explicitly training the channel over a non-explicit training scheme is a bandwidth penalty, albeit at a significantly less stringent coherence requirement on the channel.**

I. Introduction

Consider a MIMO fading channel with equal number N of transmit and receive antennas. The seminal work [1] showed that MIMO capacity, $C(N)$, scales linearly with N in a channel with independent and identically distributed (i.i.d.) Rayleigh fading between antenna pairs. This is a direct consequence of the fact that the channel power, $\rho_c(N) = \mathbf{E}\left[\text{Trace}(\mathbf{H}\mathbf{H}^H)\right] = N^2$, scales quadratically with the number of antennas. This channel power normalization is prevalent in virtually all subsequent works on MIMO channel capacity. In particular, studies based on the Kronecker product correlation model [2] show that MIMO capacity scales linearly even in correlated MIMO channels, albeit with a smaller slope. Again, $\rho_c(N)$ scales as $\mathcal{O}(N^2)$ in this case.

In a recent work [3], we argued that such quadratic scaling in channel power is not physically possible indefinitely as it would imply a linear scaling of received signal power for a given transmit power. We thus addressed the question of capacity scaling of realistic physical MIMO channels in which the channel power scales sub-quadratically with N. Specifically, the work in [3] shows that for a given channel power scaling $\rho_c(N) = \mathcal{O}(N^\gamma),\ \gamma \in (0,2]$, the capacity cannot scale faster than $C(N) = \mathcal{O}(N^{\frac{\gamma}{2}})$ and such scaling is achievable by an *ideal MIMO channel*. These results generalize our earlier results in [4] which showed that linear capacity scaling is possible for $\rho_c(N) \sim \mathcal{O}(N^2)$ whereas $C(N)$ saturates or grows at a sub-linear rate for $\rho_c(N) \sim \mathcal{O}(N)$.

Our analysis in [3], [4] is based on a unitarily equivalent virtual MIMO channel representation [5] that provides an intuitive and tractable characterization of physical MIMO channels corresponding to uniform linear arrays (ULAs) of antennas. The entries of the virtual channel matrix $\mathbf{H}_v$ are uncorrelated for any scattering environment and the correlation in $\mathbf{H}$ is captured by the *sparseness* of the dominant non-vanishing entries in $\mathbf{H}_v$. Under mild assumptions, for a given $\rho_c(N) < \mathcal{O}(N^2)$, the number of non-vanishing entries in $\mathbf{H}_v$, $D(N)$ (degrees of freedom (DoF)), also scale as $\mathcal{O}(\rho_c(N))$ compared to the maximum of $\mathcal{O}(N^2)$ in i.i.d. channels. The workhorse of [3] is a family of channels with the same channel power, but different spatial distributions of this channel power in the spatial DoF, whose capacity can be approximated as

$$C(N) \approx p(N)\log\left(1+\rho_{rx}(N)\right) = p(N)\log\left(1+\rho\frac{q(N)}{p(N)}\right) \quad (1)$$

where $p(N)$ denotes the number of parallel channels (multiplexing gain), $q(N) = D(N)/p(N)$ denotes the DoF per parallel channel, ρ is the transmit SNR and $\rho_{rx}(N) = \rho q(N)/p(N)$ is the receive SNR per parallel channel. On one extreme is the *beamforming channel*, $\mathbf{H}_{bf}$, which maximizes $\rho_{rx}(N)$ at the expense of $p(N)$ and on the other extreme is the *multiplexing channel*, $\mathbf{H}_{mux}$, which maximizes $p(N)$. The *ideal channel*, $\mathbf{H}_{id}$, lies in between with an optimal distribution of the channel power to balance $p(N)$ with $\rho_{rx}(N)$.

In addition to capacity scaling, the results in [3] also provide insights into optimal signaling as a function of SNR (ρ) for a given N and $D(N)$. There exist SNR's $\rho_{min}(N)$ and $\rho_{max}(N)$ such that $\mathbf{H}_{bf}$ yields the highest capacity for $\rho < \rho_{min}(N)$ and $\mathbf{H}_{mux}$ has the highest capacity for $\rho > \rho_{max}(N)$. Finally, an *adaptive resolution signaling framework* is proposed in [3] to create different MIMO channel configurations for any given physical scattering environment with a given number of randomly distributed (in space) propagation paths.[1]

In this paper, we build on the results in [3] as well as [6] and [7] to further investigate capacity-achieving MIMO signaling strategies in the *low-power* regime. It can be shown that the beamforming channel achieves the highest capacity and is spectrally the most efficient in the low SNR regime. We note that the beamforming channel created via adaptive resolution signaling is different from the usual notions of beamforming: essentially, all N antennas are packed in a small aperture (on the order of a wavelength), such as in a reconfigurable array, and coherently transmit the common data stream to achieve a wide beamwidth and a power amplification gain of N. Conventional capacity-optimal signaling corresponds to

This work was supported in part by the NSF under grants CCR-0113385 and CCF-0431088.

[1] The number of resolvable propagation paths equals the DoF in the MIMO channel under mild assumptions [3].

$\mathbf{H}_{mux}$ whose capacity is an order-of-magnitude smaller than that of $\mathbf{H}_{bf}$ for the chosen parameters.

In his seminal work [6], Verdu has argued that the minimum energy per bit necessary for reliable communication, $\frac{E_b}{N_o}_{\min}$, and the wideband slope, S_0, are the most important figures of merit for characterizing spectral efficiency in the low-power regime. A signaling scheme that achieves $\frac{E_b}{N_o}_{\min}$ is termed first-order optimal and one that achieves S_0 as well is termed second-order optimal. It is shown in [6] that when complete channel state information (CSI) is available at the receiver (coherent setting), QPSK signaling achieves both the first- and second-order optimality conditions for a MIMO channel. If no CSI is available at the receiver (non-coherent setting), it is shown that *flashy* signaling is necessary and sufficient to achieve the first-order optimality condition. However, a flashy signaling scheme, besides having a peak-to-average ratio that tends to ∞ (and hence practically unrealizable), also results in the second derivative of capacity converging to $-\infty$ at zero SNR, resulting in $S_0 = 0$.

In a recent work [7], Zheng *et al.* have argued that the requirement of flashy signaling is a direct consequence of the assumption that the receiver has *absolutely* no CSI. They show that in the single-antenna case, depending on the scaling of the *coherence time* with SNR, channel learning (via training symbols) coupled with a non-flashy signaling scheme can achieve capacity that is intermediate between the coherent and the non-coherent extremes.

In the next section, we present the system model and describe the family of channels introduced in [3]. The main results are presented in Sec. III which extend the results of [7] to the MIMO setting. Our results show that the coherence requirement to achieve rates that are within a constant factor of the first and second order optimality conditions are significantly less stringent for the multi-antenna channel than for achieving the exact first and second order optimality conditions. More precisely, we show that the coherence time of a beamforming channel has to satisfy $T(\mathsf{SNR}) = \frac{kp(N)}{\mathsf{SNR}^{\nu}}$ for some $k > 0$ and $\nu \geq 3$ or $\nu = 1$ for order optimality, while the requirement for precise optimality with an explicit training scheme is $T(\mathsf{SNR}) = \frac{kp(N)}{\mathsf{SNR}^{\nu}}$ with $\nu \geq 3$. The requirements for order optimality with a non-explicit training scheme is $T(\mathsf{SNR}) = \frac{kp(N)}{\mathsf{SNR}^{\nu}}$ with $\nu \geq 2$. Thus we characterize the trade-off between the bandwidth penalty and the coherence requirements on the channel to sustain rates close to the wideband capacity limit with finite bandwidth. A note about notation: $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$ means $0 < K_1 \leq \frac{f(x)}{g(x)} \leq K_2 < \infty$.

II. MIMO Channel Models

Consider a single-user MIMO system with (ULAs) of N transmit and receive antennas

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{H}$ is the MIMO channel matrix and $\mathbf{n}$ is the AWGN at the receiver. Initial results assumed an idealized channel model in which the entries of $\mathbf{H}$ are i.i.d. Gaussian random variables, reflecting a rich scattering environment. Parametric *physical* model, on the other hand, models the channel more accurately but remains mathematically intractable. The recently introduced *virtual MIMO channel representation* [5] for ULAs captures the essence of physical modeling while retaining the mathematical tractability of statistical models. The virtual channel representation essentially samples the physical scattering environment at fixed, virtual angles and is a unitarily equivalent transformation of $\mathbf{H}$

$$\mathbf{H} = \mathbf{A}_R \mathbf{H}_v \mathbf{A}_T^H \tag{3}$$

where $\mathbf{A}_R$ and $\mathbf{A}_T$ are unitary discrete Fourier transform (DFT) matrices, and $\mathbf{H}_v$ is the virtual channel matrix. The virtual channel coefficients represent the coupling between transmit and receive spatial beams in the direction of fixed virtual angles. The most important property of the virtual representation is that the virtual entries are approximately uncorrelated and this approximation improves with the number of antennas. The readers are referred to [5], [4], [3] for a more detailed description. We will work in the virtual domain and henceforth use $\mathbf{H}$ to denote $\mathbf{H}_v$.

For any given N, the $D(N)$ (independent) dominant non-zero random variables of $\mathbf{H}$ represent the statistically independent DoF in the MIMO channel. For simplicity, we assume that all the non-zero virtual entries have unit power. Thus $\rho_c(N) = D(N)$ and the channel power scaling also reflects the growth rate of the DoF in $\mathbf{H}$. We note that the i.i.d. model assumes that $\rho_c(N) = D(N) = N^2$. In this paper, we consider realistic correlated channels for which the virtual matrix is sparse[2]: $\rho_c(N) = D(N) = N^{\gamma}$, $\gamma \in (0, 2]$. Under our assumptions, it is convenient to model a sparse $N \times N$ virtual matrix with $D(N)$ DoF as $\mathbf{H} = \mathbf{M} \odot \mathbf{H}_{iid}$ where $\odot$ denotes elementwise product, $\mathbf{H}_{iid}$ is an i.i.d. matrix of $\mathcal{CN}(0, 1)$ entries, and $\mathbf{M}$ is a mask matrix with $D(N)$ unit entries and zeros elsewhere. In particular, we consider the following family of channels [3].

Definition 1: **A family of mask matrices.** Consider a mask matrix $\mathbf{M}$ with $D(N)$ non-zero entries distributed over $1 \leq p(N) \leq N$ columns and $1 \leq q(N) \leq N$ non-zero entries in each column. Let $r(N) = \max(p(N), q(N))$. Specifically, $\mathbf{M}$ is a $r(N) \times p(N)$ matrix with non-zero entries given by

$$\begin{aligned} \mathbf{M}((n+m) \mod r(N), n) &= 1 \\ n = 1 \ \cdots \ p(N), \qquad & q_-(N) \leq m \leq q_+(N) \end{aligned} \tag{4}$$

where $q_-(N) = \lceil -(q(N)-1)/2 \rceil$, $q_+(N) = \lceil (q(N)-1)/2 \rceil$, and $p(N)$ and $q(N)$ satisfy $D(N) = p(N)q(N)$. □

Note that for $q(N) \geq p(N)$, $\mathbf{M}$ is a $q(N) \times p(N)$ matrix of ones, whereas for $q(N) < p(N)$ $\mathbf{M}$ is a $p(N) \times p(N)$ matrix with $q(N)$ non-vanishing diagonals (a $q(N)$-connected $p(N)$-dimensional matrix in [4]). In all cases, there are $q(N)$ non-zero entries in each column and $p(N)$ non-zero entries in each row. Consider a given $\rho_c(N) = D(N) = N^{\gamma}$. We can generate different channels by varying $p(N) = N^{\alpha}$, $\alpha \in (0, 1)$, under the constraint $D(N) = p(N)q(N)$. The **beamforming channel**, $\mathbf{H}_{bf}$, is a $q(N) \times p(N)$ matrix with i.i.d. entries where $q(N) = N^{\min(1,\gamma)} = q_{max}(N)$ and all the DoF are distributed to maximize $q(N)/p(N)$ (minimize $p(N)$). The **multiplexing channel**, $\mathbf{H}_{mux}$, is a $q(N)$-connected $p(N) \times p(N)$ matrix

[2]Sparseness of physical MIMO channels has been experimentally observed in several recent studies.

[4] with i.i.d. non-zero entries where $q(N) = D(N)/p(N)$ and $p(N) = N^{\min(1,\gamma)} = p_{max}(N)$. The DoF are distributed to maximize the multiplexing gain $p(N)$. The **ideal channel**, $\mathbf{H}_{id}$, is between the two extremes and is a $\sqrt{D(N)} \times \sqrt{D(N)}$ matrix with i.i.d. entries: the DoF are distributed so that $p(N) = q(N) = \sqrt{D(N)}$. A pictorial illustration of the three channels can be found in [3] where we have also shown that the beamforming channel achieves the highest capacity and the smallest possible $\frac{E_b}{N_o}_{\min}$ at low SNR. Even though the beamforming channel does not achieve the largest S_0 value for the three channels, the operating point in the spectral efficiency vs. $\frac{E_b}{N_o}$ curve is determined primarily by $\frac{E_b}{N_o}_{\min}$. Therefore in this paper, we consider the beamforming channel under low SNR conditions and study the coherence requirements on the channel to achieve near coherent capacity performance.

III. Non-flashy Signaling in Beamforming Channels

A. Description of Training and Communication Scheme

We consider a non-flashy channel training and communication scheme and we use $\dot{C}_{NF}(0)$ and $\ddot{C}_{NF}(0)$ to denote respectively the first and the second derivatives of the average mutual information of this training scheme at zero SNR. We study the coherence cost of the channel so that $\dot{C}_{NF}(0) = \mathcal{O}\left(\dot{C}_{bf}(0)\right)$ and $\ddot{C}_{NF}(0) = \mathcal{O}\left(\ddot{C}_{bf}(0)\right)$. We now describe the family of non-flashy communication schemes over one coherence length T. This scheme is a slight modification of the training scheme used for the single antenna case in [7].

The transmitter sends known symbols during a training phase of length T_{tr} which helps the receiver in making an estimate of the channel. The transmitter communicates over the remaining $T - T_{tr}$ symbol periods. The total energy E_{tot} used for communication and training is normalized to T SNR. A fraction η of this energy is used in training and is denoted by E_{tr} while the remaining $(1-\eta)$ fraction of E_{tot} is used for communication. For notational simplicity, we will denote by P the energy used over one training symbol, i.e. $P = \frac{E_{tr}}{T_{tr}} = \frac{\eta T \text{ SNR}}{T_{tr}}$.

Let $\mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_{T_{tr}}]$ denote the matrix of collective training symbols. Since the transmitter has no CSI during the training phase, $\mathbf{S}$ can be written as $\mathbf{S} = \sqrt{\frac{T_{tr}P}{p(N)}}\mathrm{U}$ with $\mathrm{U}\mathrm{U}^H = \mathrm{I}_{p(N)}$. The received vector $\mathrm{y_t}$ at time t of the training period is given by

$$\mathrm{y_t} = \sqrt{\frac{T_{tr}P}{p(N)}}\mathbf{H}\mathrm{U_t} + \mathbf{n}_\mathrm{t} \tag{5}$$

where $\mathrm{U_t}$ is the transmitted data during the t-th training period and $\mathbf{n}_\mathrm{t}$ is the AWGN. Let $\mathrm{Y} = [\mathrm{y}_1 \cdots \mathrm{y}_{T_{tr}}]$ and $\mathbf{N} = [\mathbf{n}_1 \cdots \mathbf{n}_{T_{tr}}]$. We then have

$$\mathrm{Y} = \sqrt{\frac{T_{tr}P}{p(N)}}\mathbf{H}\mathrm{U} + \mathbf{N}. \tag{6}$$

The following is well-known from estimation theory. The MMSE estimate $\hat{\mathbf{H}}$ of $\mathbf{H}$ is given by $\hat{\mathbf{H}} = \frac{\sqrt{\frac{T_{tr}P}{p(N)}}}{1+\frac{T_{tr}P}{p(N)}}\mathrm{Y}\mathrm{U}^H$. Since the channels under consideration in this paper are regular, the estimation errors of the channel coefficient $\mathbf{H}_{ij}$ are identical and is given by $\sigma^2 = \frac{1}{1+\frac{T_{tr}P}{p(N)}}$. We have the following proposition on the estimation error which would be useful in computing the lower bound to capacity in Theorem 1.

Proposition 1: Let $\mathbf{H}$ be a regular multi-antenna channel, $\hat{\mathbf{H}}$ be its MMSE estimate, and $\mathbf{x}$ be the $p(N)$-dimensional transmitted vector with covariance matrix $\mathbf{Q} = \frac{\mathrm{Tr}(\mathbf{Q})}{p(N)}\mathrm{I}_{p(N)}$. If $\Delta = \mathbf{H} - \hat{\mathbf{H}}$ is the estimation error matrix, then the covariance matrix $\mathbf{\Sigma}_{\Delta\mathbf{x}} = \mathrm{Tr}(\mathbf{Q})\sigma^2\mathrm{I}_{q(N)}$.

Since the quality of the MMSE estimate of $\mathbf{H}$ does not depend on the training time T_{tr} as long as $T_{tr} \geq p(N)$ and only depends on E_{tr} [9], we can assume without loss of generality that $T_{tr} = p(N)$.

B. Conditions for Optimality of Non-Flashy Scheme

We are now prepared to state the main statement of this paper.

Theorem 1: Let $\mathbf{H}$ be a $q(N) \times p(N)$ beamforming channel. Consider a non-flashy training and communication scheme as described above. Then $\dot{C}_{NF}(0)$ and $\ddot{C}_{NF}(0)$ are such that $\dot{C}_{NF}(0) = \mathcal{O}\left(\dot{C}_{bf}(0)\right)$ and $\ddot{C}_{NF}(0) = \mathcal{O}\left(\ddot{C}_{bf}(0)\right)$ if and only if the coherence length satisfies $T \geq k\frac{p(N)}{\text{SNR}^\nu}$ for $\nu = 1$ or $\nu \geq 3$ and for some $k > 0$.

Proof: A lower bound to C_{bf} is the average mutual information I_{lower} of the training based communication scheme with a Gaussian input $\mathbf{x}$ having a transmit covariance matrix $\mathbf{Q} = \frac{\mathrm{Tr}(\mathbf{Q})}{p(N)}\mathrm{I}_{p(N)}$ where $\mathrm{Tr}(\mathbf{Q}) = \frac{E_{tot}-E_{tr}}{T-T_{tr}} = \frac{(1-\eta)T \text{ SNR}}{T-T_{tr}}$. This average mutual information can be lower bounded using techniques similar to [10] to yield

$$C_{bf} \geq \frac{T-T_{tr}}{T}\mathbf{E}\left[\log_2\det\left(\mathrm{I} + \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H\left(\mathrm{I}+\mathbf{\Sigma}_{\Delta\mathbf{x}}\right)^{-1}\right)\right].$$

Using Proposition 1 to rewrite $\mathbf{\Sigma}_{\Delta\mathbf{x}}$, we can further lower bound the capacity by

$$C_{bf} \geq \frac{T-T_{tr}}{T}\mathbf{E}\left[\log_2\det\left(\mathrm{I} + \mu\hat{\mathbf{H}}\hat{\mathbf{H}}^H\right)\right] \tag{7}$$

where $\mu = \frac{\mathrm{Tr}(\mathbf{Q})}{p(N)(1+\mathrm{Tr}(\mathbf{Q})\sigma^2)}$. The channel estimate $\hat{\mathbf{H}}$ is another beamforming channel with variance of entries being $1-\sigma^2$. Following the approximate capacity expression (1), we can compute this lower bound I_{lower} as

$$\mathrm{I}_{lower} \approx \frac{T-T_{tr}}{T}p(N)\log_2\left(1+\mu q(N)(1-\sigma^2)\right) \tag{8}$$

and after simplification we have

$$C_{bf} \geq \frac{T-T_{tr}}{T}p(N)\log_2\left(1+\frac{\text{SNR}^2\ T^2 q(N)\eta(1-\eta)}{p(N)a}\right) \tag{9}$$

where $a = p(N)(T - T_{tr} + (1-\eta)T\text{ SNR}) + \eta T(T - T_{tr})\text{ SNR}$. Maximizing the above expression as a function of η, we obtain $\eta_{\max} = \frac{p(N)(T-T_{tr}+T\text{ SNR})}{T\text{ SNR}(T-T_{tr}-p(N))}\left(\sqrt{1+\frac{T(T-T_{tr}-p(N))\text{ SNR}}{p(N)(T-T_{tr}+T\text{ SNR})}}-1\right)$, which results in

$$\begin{aligned} C_{bf} &\geq \frac{T-T_{tr}}{T}p(N)\log_2\left(1+\frac{q(N)\ b}{p(N)(T-T_{tr}-p(N))^2}\right) \\ &\geq \frac{T-T_{tr}}{T}p(N)\log_2\left(1+\frac{q(N)c(T-T_{tr})}{p(N)(T-T_{tr})^2}\right) \\ &= \frac{T-T_{tr}}{T}p(N)\log_2\left(1+\frac{q(N)c}{p(N)(T-T_{tr})}\right) \end{aligned} \tag{10}$$

where $b = T\ \mathsf{SNR}(T-T_{tr}+p(N)) + 2p(N)(T-T_{tr}) - 2\sqrt{p(N)(p(N)+T\ \mathsf{SNR})(T-T_{tr})(T-T_{tr}+T\ \mathsf{SNR})}$ and $c = 2p(N) - 2\sqrt{p(N)(p(N)+T\ \mathsf{SNR})\left(1+\frac{T\ \mathsf{SNR}}{T-T_{tr}}\right)} + T\ \mathsf{SNR}$. The proof is complete following the discussion in the next two sub-sections. □

C. Characterization of Low-SNR Asymptotics

In this sub-section, we characterize the exact low-SNR asymptotics that $\frac{T-T_{tr}}{T}$ and $\frac{c}{T-T_{tr}}$ of (10) have to satisfy so that $\dot{C}_{NF}(0) = \mathcal{O}\left(\dot{C}_{bf}(0)\right)$ and $\ddot{C}_{NF}(0) = \mathcal{O}\left(\ddot{C}_{bf}(0)\right)$. Henceforth all constants of the form k_i, m_i and n_i will be assumed to be finite unless specified otherwise.

We assume that the low-SNR asymptotics of $\frac{T-T_{tr}}{T}$ satisfies $\frac{T-T_{tr}}{T} = k_1 - k_2\mathsf{SNR}^{\alpha}$ where $\alpha \geq 0$. To satisfy the positivity of this fraction, we would need $0 < k_1 - k_2 \leq 1$ if $\alpha = 0$, or $k_2 > 0$ and $0 < k_1 \leq 1$ if $\alpha > 0$. We assume that the low-SNR asymptotics of $\frac{c}{T-T_{tr}}$ is $k_3\mathsf{SNR}^{\beta} + k_4\mathsf{SNR}^{\gamma}$ where $\gamma\ (> \beta > 0)$ is the index of the second-dominant term of the Taylor's series expansion of $\frac{c}{T-T_{tr}}$ and $k_3 \neq 0$. Expanding the Taylor's series of $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3}$ and ignoring the higher order terms (h.o.t.), we can write the lower-bound to C_{bf} in (10) as

$$\begin{aligned}\frac{C_{bf}}{\log_2(e)} &\geq q(N)\left(k_1 - k_2\mathsf{SNR}^{\alpha}\right)\left(k_3\mathsf{SNR}^{\beta} + k_4\mathsf{SNR}^{\gamma}\right)\\ &- \frac{q(N)^2}{2p(N)}\left(k_1 - k_2\mathsf{SNR}^{\alpha}\right)\left(k_3\mathsf{SNR}^{\beta} + k_4\mathsf{SNR}^{\gamma}\right)^2\\ &+ \text{h.o.t.}\end{aligned} \tag{11}$$

We note that if $\beta \in (0,1)$ then $\dot{C}_{NF}(0) = \infty$, which would mean that $\frac{E_b}{N_o}_{\min} = 0$, which we know is not the case. If $\beta > 1$, then $\dot{C}_{NF}(0) = 0$ which violates the first order optimality condition. Thus $\beta = 1$ and $\gamma > 1$. From (11), if either $\gamma < 2$ or $\alpha \in (0,1)$ then the second order optimality condition is violated since terms of the form SNR^{δ} where $\delta \in (1,2)$ arise. Thus $\gamma \geq 2$ and either $\alpha = 0$ or $\alpha \geq 1$.

Thus the following are the only possible choices of α, β and γ for both first and second order optimality: $\alpha = 0, \beta = 1, \gamma \geq 2$,[3] and $\alpha \geq 1, \beta = 1, \gamma \geq 2$.[4] We note that in the second case γ is independent of the choice of α. Recasting the above characterizations: the low-SNR asymptotics of $\frac{T-T_{tr}}{T}$ and $\frac{c}{T-T_{tr}}$ has to satisfy either

$$\begin{aligned}\text{Case 1}: \quad & \frac{T-T_{tr}}{T} = m_1, & 0 < m_1 \leq 1\\ & \frac{c}{T-T_{tr}} = m_2\mathsf{SNR} + m_3\mathsf{SNR}^{\gamma}, & \gamma \geq 2,\\ \text{Case 2}: \quad & \frac{T-T_{tr}}{T} = m_4 - m_5\mathsf{SNR}^{\alpha} & \alpha \geq 1\\ & \frac{c}{T-T_{tr}} = m_6\mathsf{SNR} + m_7\mathsf{SNR}^{\gamma}, & \gamma \geq 2, m_6 > 0.\end{aligned}$$

No further terms are needed in the Taylor's series expansion of $\frac{c}{T-T_{tr}}$ of the form $k_3\mathsf{SNR}^{\beta} + k_4\mathsf{SNR}^{\gamma} + k_5\mathsf{SNR}^{\epsilon}$ since $\epsilon > 2$ and even the x term in the Taylor's expansion of $\log(1+x)$ would not have a contribution from $k_5\mathsf{SNR}^{\epsilon}$ in either the first or second derivative. Similarly, the expansion of $\frac{T-T_{tr}}{T}$ need not be of the form $k_1 - k_2\mathsf{SNR}^{\alpha} + k_6\mathsf{SNR}^{\mu}$ since $\mu > 1$ and $\beta = 1$ would mean that there is no contribution from $k_6\mathsf{SNR}^{\mu}$ in the first or the second derivatives of the lower bound to C_{bf}.

[3] We do not need additional constraints on the k_i in this case since $k_1 - k_2 > 0$ by assumption.

[4] We need the additional condition that $k_3 > 0$ for this case.

D. Characterization of the Coherence Time

In this section, we answer the question: Does there exist a coherence time $T = T(\mathsf{SNR})$ such that either of Case 1 or 2 characterized in the previous section can be satisfied? Consider Case 1 first. If $\frac{T-T_{tr}}{T} = m_1$, we can write $\frac{c}{T-T_{tr}}$ as

$$\begin{aligned}\frac{c}{T-T_{tr}} &= \frac{c}{T}\frac{T}{T-T_{tr}} = \frac{c}{m_1 T} = \frac{\mathsf{SNR}}{m_1 f(\mathsf{SNR})}g(\mathsf{SNR})\\ &= \frac{\mathsf{SNR}}{m_1 f(\mathsf{SNR})} + \frac{\mathsf{SNR}}{2m_1} -\\ &\quad \frac{\mathsf{SNR}}{m_1 f(\mathsf{SNR})}\sqrt{1+f(\mathsf{SNR})}\sqrt{1+\frac{\mathsf{SNR}}{m_1}}\end{aligned}$$

where $g(\mathsf{SNR}) = 1 + \frac{f(\mathsf{SNR})}{2} - \sqrt{(1+f(\mathsf{SNR}))(1+\frac{\mathsf{SNR}}{m_1})}$ and $f(\mathsf{SNR}) = \frac{T\mathsf{SNR}}{p(N)}$. Given that $\frac{T-T_{tr}}{T} = m_1$, the function $f(\mathsf{SNR})$ satisfies $f(\mathsf{SNR}) = n_1\mathsf{SNR}$, $n_1 > 0$. This implies that $\frac{c}{T-T_{tr}} \to n_2$, $n_2 \neq 0$ which contradicts the fact that $\lim_{\mathsf{SNR}\to 0}\frac{c}{T-T_{tr}} = 0$. Thus Case 1 cannot be satisfied by any choice of $T(\mathsf{SNR})$.

In Case 2, we can write $\frac{c}{T-T_{tr}}$ after simplification as

$$\begin{aligned}\frac{c}{T-T_{tr}} = \frac{2}{m_4}\left(\frac{\mathsf{SNR}}{f(\mathsf{SNR})} + \frac{m_5\mathsf{SNR}^{1+\alpha}}{m_4 f(\mathsf{SNR})}\right) \times \left(1 + \frac{f(\mathsf{SNR})}{2}\right.\\ \left. - \sqrt{1+f(\mathsf{SNR})}\left(1 + \frac{\mathsf{SNR}}{2m_4} + \frac{m_5\mathsf{SNR}^{1+\alpha}}{2m_4^2}\right)\right).\end{aligned}$$

We consider $f(\mathsf{SNR}) = \mathsf{SNR}^{\eta}$ for various values of η. The first condition of Case 2 then implies that $\eta = 1 - \alpha$ and $T = \frac{p(N)}{\mathsf{SNR}^{\alpha}}$. We ignore constants preceding SNR-powers to keep the algebra simple. We get the following simplification for $\frac{c}{T-T_{tr}}$:

$$\begin{aligned}\frac{c}{T-T_{tr}} &= \mathsf{SNR} + \mathsf{SNR}^{\alpha} + \mathsf{SNR}^{2\alpha} + \mathsf{SNR}^{1+\alpha}\\ &\quad + \sqrt{1+\mathsf{SNR}^{1-\alpha}}\cdot\Big(\mathsf{SNR}^{\alpha} + \mathsf{SNR}^{2\alpha}\\ &\quad + \mathsf{SNR}^{1+\alpha} + \mathsf{SNR}^{2\alpha+1} + \mathsf{SNR}^{3\alpha+1}\Big).\end{aligned} \tag{12}$$

The term $\sqrt{1+\mathsf{SNR}^{1-\alpha}}$ can be expanded for $\alpha > 1$ as $\mathsf{SNR}^{\frac{1-\alpha}{2}} + \mathsf{SNR}^{\frac{\alpha-1}{2}}$ which reduces $\frac{c}{T-T_{tr}}$ to

$$\begin{aligned}\frac{c}{T-T_{tr}} &= \mathsf{SNR} + \mathsf{SNR}^{\alpha} + \mathsf{SNR}^{2\alpha} + \mathsf{SNR}^{1+\alpha} +\\ &\quad \mathsf{SNR}^{\frac{1+\alpha}{2}} + \mathsf{SNR}^{\frac{1+3\alpha}{2}} + \mathsf{SNR}^{\frac{3+\alpha}{2}} +\\ &\quad \mathsf{SNR}^{\frac{3\alpha+3}{2}} + \mathsf{SNR}^{\frac{3+5\alpha}{2}} + \mathsf{SNR}^{\frac{3\alpha-1}{2}} +\\ &\quad \mathsf{SNR}^{\frac{5\alpha-1}{2}} + \mathsf{SNR}^{\frac{1+5\alpha}{2}} + \mathsf{SNR}^{\frac{1+7\alpha}{2}}.\end{aligned} \tag{13}$$

If $\alpha \in (1,3)$, second order optimality is violated by atleast one of the $\mathsf{SNR}^{\alpha}, \mathsf{SNR}^{\frac{1+\alpha}{2}}$ and $\mathsf{SNR}^{\frac{3\alpha-1}{2}}$ terms. Second order optimality holds precisely when $\alpha \geq 3$. The only other value of α not considered thus far is $\alpha = 1$ when the expansion in (13) does not hold. However from (12) we see that every power of SNR is either 1 or greater than equal to 2. Thus we have both first and second order optimality in this case. Thus first and second order order-optimality are satisfied precisely when $T = k\frac{p(N)}{\mathsf{SNR}^{\alpha}}$ with $\alpha = 1$ or $\alpha \geq 3$ and for some $k > 0$.

IV. Discussion

In the SISO case, where $p(N) = 1$, Zheng *et al.* showed that the coherence time has to satisfy $T = \frac{k}{\mathsf{SNR}^\nu}$ for some $\nu \geq 1+2\epsilon$ and $k > 0$ for a non-flashy training-based signalling scheme to achieve capacity of the form $\mathsf{SNR} - \mathcal{O}(\mathsf{SNR}^{1+\epsilon})$. In particular, for the coherent capacity with $\epsilon = 1$, the coherence time has to satisfy $T \geq \frac{k}{\mathsf{SNR}^3}$. Theorem 1 shows that with training-based schemes, rates close to coherent capacity can not only be achieved with $\nu \geq 3$ but also with a less constrained coherence time corresponding to $\nu = 1$. Zheng *et al.* also study the performance of joint training and communication schemes that do not have explicit training slots and show that to achieve capacity of the form $\mathsf{SNR} - \mathcal{O}(\mathsf{SNR}^{1+\epsilon})$ the coherence time has to satisfy $T = \frac{k}{\mathsf{SNR}^\nu}$ for some $\nu \geq 2\epsilon$ and $k > 0$. Extension to the MIMO case is reported in [8].

Comparing the coherence cost of the two schemes, [7] and [8] claim that a training-based communication scheme is strictly sub-optimal in achieving coherent capacity when compared with a scheme that does no explicit training. However by studying schemes that achieve rates of the form $\mathcal{O}(\mathsf{SNR}) - \mathcal{O}(\mathsf{SNR}^2)$ in the SISO case (and corresponding MIMO rates), we have shown that training-based schemes could be optimal over non-training schemes in terms of coherence cost. To expound more on this, we study the bandwidth penalty of three schemes: training-based scheme S_1 for a channel with $T = \frac{k_1 p(N)}{\mathsf{SNR}^\nu},\ \nu > 3$, joint training and communication scheme S_2 for a channel with $T = \frac{k_2 p(N)}{\mathsf{SNR}^\nu},\ \nu > 2$, and training-based scheme S_3 for a channel with $T = \frac{k_3 p(N)}{\mathsf{SNR}}$.

The first and second derivatives of the average mutual information of scheme S_1 at zero SNR is given by $\dot{C}_{S_1}(0) = q(N)\log_2(e)$ and $\ddot{C}_{S_1}(0) = -\frac{q(N)^2}{2p(N)}\log_2(e)$. The first and second derivatives of average mutual information of S_2 are the same as that of S_1. However with scheme S_3 it is easy to check that $\dot{C}_{S_3}(0) = \log_2(e)q(N) \cdot \frac{\left(\sqrt{1+k_3}-1\right)^2}{k_3} \stackrel{\leq}{\to} \log_2(e)q(N)$ and $\frac{\ddot{C}_{S_3}(0)}{\log_2(e)} = -\left(\frac{q(N)^2}{2p(N)} \cdot \frac{\left(\sqrt{1+k_3}-1\right)^4}{k_3^2} + q(N) \cdot \frac{\sqrt{1+k_3}}{k_3}\right) \to -\frac{q(N)^2}{2p(N)}$ as $k_3 \to \infty$.

This analysis shows that both S_1 and S_2 achieve the minimum energy per bit needed for reliable communication and wideband slope, that is equal to the corresponding quantities assuming perfect CSI at the receiver. This also reflects in the bandwidth penalty that has to be paid to achieve a rate R at a power $P > P_{min} \doteq N_0 R \frac{E_b}{N_o}_{\min}$. With power $P = P_{min}$, an infinite bandwidth is required to sustain this rate. However the same rate R can be sustained with a finite bandwidth $B \approx \frac{R^2}{P-P_{min}} \frac{N_0 \log_e(2)}{S_0} \frac{E_b}{N_o}_{\min}$ provided that the Taylor's expansion is valid at P [6]. Since the two schemes S_1 and S_2 have the same $\frac{E_b}{N_o}_{\min}$ and S_0, the bandwidths of the two schemes can be compared and it can be seen that both the schemes require the same bandwidth to achieve a rate R at a power level P. However the coherence cost of S_2 is much lower than that of S_1, as pointed out by [7].

The $\frac{E_b}{N_o}_{\min}$ values of S_1 and S_3 differ by a constant and a bandwidth-rate relationship as above cannot be used to compare the two schemes. However the concavity of the average mutual information as a function of SNR can be used to provide a lower bound on the bandwidth required by S_1 to achieve rate R at power level $P > P_{min}\big|_{S_3}$. Thus the bandwidths required by the two schemes for sufficiently large $q(N)$ can be bounded as follows: $\frac{B\big|_{S_3}}{B\big|_{S_1}} \leq \frac{k_3}{\left(\sqrt{1+k_3}-1\right)^2}$. As the constant k_3 increases, the bandwidth required by S_3 converges to that of S_1 (which is the same as that for a coherent channel). More notably, when the channel has a coherence time $T = \frac{k_4 p(N)}{\mathsf{SNR}^\nu},\ \nu \in (1,3)$, the second order optimality condition fails (see Sec. III D). This implies that the training scheme achieves a zero wideband slope and we *cannot* sustain any rate R, however small, with finite bandwidth. For the joint training and communication scheme, no rate R can be sustained with a finite bandwidth if the coherence time satisfies $T = \frac{kp(N)}{\mathsf{SNR}^\nu}$ for some $\nu < 2$ and $k > 0$. Thus we have shown that by trading off the bandwidth penalty, rates close to the wideband capacity limit can be sustained at finite bandwidths by training-based schemes for significantly lower coherence constraint on the channel. Such a trade-off is *impossible* with joint training and communication schemes.

Another advantage of the training-based scheme is illustrated below. For a training scheme to achieve optimality, the training energy E_{tr} should be such that the channel can be learnt reliably and the fraction of the total energy used for training, η, should be negligible [7]. It is shown in [7] that $\eta \leq \mathsf{SNR}$ and $\frac{E_{tr}}{T_{tr}} \geq \frac{1}{\mathsf{SNR}}$ are necessary to attain first and second order optimality in the SISO case. Thus the total transmitted energy per coherent block $E_{tot} = T\mathsf{SNR} \geq \frac{k}{\mathsf{SNR}^2}$ with the training scheme and $E_{tot} \geq \frac{k}{\mathsf{SNR}}$ with the joint training and communication scheme. By focussing on achieving order optimality, that is, by letting $\frac{E_b}{N_o}_{\min}$ and S_0 differ from their coherent values by a constant, we have shown that the total energy per coherent block can remain finite and still lead to reliable communication. This is extremely crucial in constructing practical peak-constrained RF chains which are rendered impossible with non-training based communication schemes.

References

[1] Í. E. Telatar, "Capacity of Multi-antenna Gaussian channels," *AT& T-Bell Labs Internal Tech. Memo.,* June 1995

[2] Chen-Nee Chuah, D. N. C. Tse, J. M. Kahn, R. A. Valanzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637-650, Mar. 2002.

[3] A. M. Sayeed, V. Raghavan, J. H. Kotecha, "Capacity of space-time wireless channels: A physical perspective," *Proc of Inform. Theory Workshop*, San Antonio, TX, 2004

[4] K. Liu, V. Raghavan, A. M. Sayeed, "Capacity scaling and spectral efficiency in wideband correlated MIMO channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2504-2526, Oct. 2003

[5] A. M. Sayeed, "Deconstructing Multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002

[6] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002

[7] L. Zheng, M. Medard, D. N. C. Tse, "Channel Coherence in the Low SNR Regime," *To be submitted to IEEE Trans. Inform. Theory*, 2005.

[8] S. Ray, M. Medard, L. Zheng, "On Non-coherent MIMO Capacity in the Wideband Regime: Interplay of SNR, Antennas and Coherence," *Submitted to IEEE Trans. Inform. Theory*, Mar. 2005. *Also ISIT 2005.*

[9] B. Hassibi, B. M. Hochwald, "How much training is needed in a multiple-antenna wireless link?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-964, Apr. 2003

[10] M. Medard, "The Effect Upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 935-946, May 2000