Capacity Bounds for Peak-Constrained Multiantenna Wideband Channels

Ulrich G. Schuster, Giuseppe Durisi, Helmut Bölcskei
Communication Technology Laboratory
ETH Zurich, 8092 Zurich, Switzerland
E-mail: {schuster, gdurisi, boelcskei}@nari.ee.ethz.ch

H. Vincent Poor
Department of Electrical Engineering
Princeton University, Princeton, NJ 08544, U.S.A.
Email: poor@princeton.edu

Abstract—**This paper presents bounds on the noncoherent capacity of a very general multiple-input multiple-output channel, which allows for selectivity in time and frequency as well as for spatial correlation. The bounds apply to peak-constrained inputs; they are explicit in the channel's scattering function, are useful for a large range of bandwidth, and allow one to coarsely identify the capacity-optimal combination of bandwidth and number of transmit antennas. Furthermore, a closed-form expression is obtained for the first-order Taylor series expansion of capacity in the limit of infinite bandwidth. From this expression, it is concluded that in the wideband regime: (i) it is optimal to use only one transmit antenna when the channel is spatially uncorrelated; (ii) rank-one statistical beamforming is optimal if the channel is spatially correlated; and (iii) spatial correlation, be it at the transmitter, the receiver, or both, is beneficial.**

I. Introduction and Summary of Results

Bandwidth and space are sources of degrees of freedom that can be used to transmit information over wireless fading channels. Channel measurements indicate that an increase in the number of degrees of freedom also increases the channel uncertainty the receiver must resolve [1]. If the transmit signal is allowed to be peaky, i.e., if it can have unbounded peak power, channel uncertainty is immaterial in the limit of infinite bandwidth; indeed, for a fairly general class of fading channels, the capacity of the infinite-bandwidth additive white Gaussian noise (AWGN) channel can be achieved [2], [3]. A more realistic modeling assumption is to limit the peak power of the transmitted signal. In this case, the capacity behavior of most channels changes drastically: for certain types of peak constraint, the capacity can even approach zero in the wideband limit [3]–[5]. Intuitively, under a peak constraint on the transmit signal, the receiver is no longer able to resolve the channel uncertainty as the number of degrees of freedom increases. Consequently, issues of significant practical relevance are how much bandwidth to use, and whether spatial degrees of freedom obtained through multiple antennas can be exploited to increase capacity.

The aim of this paper is to characterize the capacity of spatially correlated multiple-input multiple-output (MIMO) fading channels that are time and frequency selective, i.e., that exhibit memory in frequency, and time, given that (i) the input signal has bounded peak power and (ii) the transmitter and the receiver know the channel law but both are ignorant of the channel realization. The assumptions (ii) constitute the *noncoherent setting*, as opposed to the *coherent setting* where the receiver has perfect channel state information (CSI) and the transmitter knows the channel law only.

Related Work: Sethuraman *et al.* [6] derived a closed-form expression for the low-SNR capacity of peak-constrained MIMO Rayleigh-fading channels that are frequency flat, time selective, and spatially uncorrelated. In particular, they show that it is optimal to use only a single transmit antenna in the low-SNR regime, and that additional receive antennas are always beneficial.

Spatial correlation is often beneficial in the noncoherent setting. Its impact on the capacity of memoryless channels has been discussed in [7], on the rates achievable with specific signaling schemes on both memoryless and block-fading channels in [8], [9], and on the reliability function at low SNR in [10].

Contributions: We consider a point-to-point MIMO channel where each *component channel* between a given transmit antenna and a given receive antenna is *underspread* and satisfies the standard *wide-sense stationary uncorrelated-scattering* (WSSUS) assumption [11]; our channel model allows for selectivity in time and frequency. We assume that the component channels are spatially correlated according to the *separable* (Kronecker) correlation model [12] and that they are characterized by the same scattering function; furthermore, the transmit signal is peak constrained. On the basis of a discrete-time, discrete-frequency approximation of said channel model that is enabled by the underspread property [13], we obtain the following results (proven in [14]):

- We present upper and lower bounds on capacity. These bounds are explicit in the channel's scattering function and allow one to coarsely identify the capacity-optimal combination of bandwidth and number of transmit antennas for a fixed number of receive antennas.
- For spatially uncorrelated channels, we generalize the asymptotic results of Sethuraman *et al.* [6] to time- *and* frequency-selective channels: for large enough bandwidth—or equivalently, for small enough SNR—it is optimal to use a single transmit antenna only, while additional receive antennas always increase capacity.
- In the wideband regime, we find that both transmit *and* receive correlation increase capacity and that rank-one

This work was supported partly by the European Commission through the Integrated Project PULSERS Phase II under contract No. FP6-027142, and partly by the U.S. National Science Foundation under Grants ANI-03-38807 and CNS-06-25637.

statistical beamforming along the strongest eigenmode of the spatial transmit correlation matrix is optimal.

Notation: Uppercase boldface letters denote matrices, and lowercase boldface letters designate vectors. The superscripts ^{T}, *, and ^{H} stand for transposition, element-wise conjugation, and Hermitian transposition, respectively. The Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$. We designate the identity matrix of dimension $N \times N$ by $\mathbf{I}_N$, the determinant of the square matrix $\mathbf{X}$ by $\det(\mathbf{X})$, and its rank by $\operatorname{rank}(\mathbf{X})$. We let $\operatorname{diag}\{\mathbf{x}\}$ denote a diagonal square matrix whose main diagonal contains the elements of the vector $\mathbf{x}$. The function $\delta(x)$ is the Dirac distribution. All logarithms are to the base e. For two functions $f(x)$ and $g(x)$, the notation $f(x) = o(g(x))$ means that $\lim_{x\to 0} f(x)/g(x) = 0$. Finally, we denote expectation by $\mathbb{E}[\cdot]$ and the Fourier transform by $\mathbb{F}[\cdot]$.

II. System Model

In the following, we first introduce the single-input single-output (SISO) model for one component channel and subsequently discuss the extension of this model to the MIMO setting.

A. Underspread WSSUS Channels

The relation between the input $x(t)$ and the corresponding output $y(t)$ of a SISO stochastic linear time-varying channel $\mathbb{H}$ can be expressed as

$$y(t) = \big(\mathbb{H}\, x\big)(t) + w(t) = \int_{t'} k_{\mathbb{H}}(t,t')x(t')dt' + w(t) \quad (1)$$

where $k_{\mathbb{H}}(t,t')$ denotes the random kernel of the channel $\mathbb{H}$ and $w(t)$ is a white Gaussian noise process. We assume that $k_{\mathbb{H}}(t,t')$ is a zero-mean jointly proper Gaussian (JPG) process in t and t' whose Fourier transforms are well defined. In particular, $L_{\mathbb{H}}(t,f) = \mathbb{F}_{\tau\to f}[k_{\mathbb{H}}(t,t-\tau)]$ is called the *time-varying transfer function* and $S_{\mathbb{H}}(\nu,\tau) = \mathbb{F}_{t\to\nu}[k_{\mathbb{H}}(t,t-\tau)]$ is called the *spreading function*. We assume that the channel is WSSUS, so that $\mathbb{E}[S_{\mathbb{H}}(\nu,\tau)S^*_{\mathbb{H}}(\nu',\tau')] = C_{\mathbb{H}}(\nu,\tau)\delta(\nu-\nu')\delta(\tau-\tau')$. Consequently, the statistical properties of the channel $\mathbb{H}$ are completely specified through its *scattering function* $C_{\mathbb{H}}(\nu,\tau)$. A WSSUS channel is called *underspread* [13] if $C_{\mathbb{H}}(\nu,\tau)$ is compactly supported on a rectangle $[-\nu_0,\nu_0]\times[-\tau_0,\tau_0]$ whose *spread* $\Delta_{\mathbb{H}} = 4\nu_0\tau_0$ satisfies $\Delta_{\mathbb{H}} < 1$. Virtually all wireless channels are highly underspread, that is, their spread satisfies $\Delta_{\mathbb{H}} \ll 1$.

B. Discrete Approximation

To simplify information-theoretic analysis, we would like to *diagonalize* the channel $\mathbb{H}$ and replace the integral input-output (IO) relation (1) by a *countable* set of *scalar* IO relations. To this end, we cannot use an eigendecomposition of the random kernel $k_{\mathbb{H}}(t,t')$, because its eigenfunctions are random as well, and hence unknown to the transmitter and the receiver in the noncoherent setting. Yet, for underspread channels it is possible to find an orthonormal set of *deterministic* approximate eigenfunctions that depend only on the channel's scattering function [13]. Consequently, knowledge of the channel law—and hence of the scattering function—is sufficient for transmitter and receiver to approximately diagonalize $\mathbb{H}$. One possible choice of approximate eigenfunctions is the *Weyl-Heisenberg set* of mutually orthogonal time-frequency shifts $g_{k,n}(t) = g(t-kT)e^{j2\pi nFt}$ of some prototype function $g(t)$ that is well localized in time and frequency. The grid parameters T and F need to satisfy $TF \geq 1$; then, the kernel of $\mathbb{H}$ can be approximated as [15]

$$k_{\mathbb{H}}(t,t') \approx \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \underbrace{L_{\mathbb{H}}(kT,nF)}_{h[k,n]} g_{k,n}(t)g^*_{k,n}(t'). \quad (2)$$

The approximation quality, as well as the choice of the prototype function $g(t)$ and the parameters T and F, are discussed in [13], [15]. The eigenvalues of the approximate channel with kernel (2) are given by $h[k,n] = L_{\mathbb{H}}(kT,nF)$. As the channel is JPG and WSSUS, the discretized channel process $\{h[k,n]\}$ is also JPG and stationary in both discrete time k and discrete frequency n. We denote its correlation function by $R[k,n] = \mathbb{E}[h[k'+k,n'+n]h^*[k',n']]$, normalized as $R[0,0]=1$. The associated spectral density

$$c(\theta,\varphi) = \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} R[k,n]e^{-j2\pi(k\theta-n\varphi)}, \quad |\theta|,|\varphi| \leq 1/2$$

can be expressed in terms of $C_{\mathbb{H}}(\nu,\tau)$ as [15]

$$c(\theta,\varphi) = \frac{1}{TF}\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta-k}{T},\frac{\varphi-n}{F}\right). \quad (3)$$

We choose $T \leq 1/(2\nu_0)$ and $F \leq 1/(2\tau_0)$ so that no aliasing of the scattering function occurs in (3). Next, we substitute the approximation (2) into (1) and project the input $x(t)$ and the output $y(t)$ onto the orthogonal set $\{g_{k,n}(t)\}$ to obtain the countable set of scalar IO relations

$$y[k,n] = h[k,n]x[k,n] + w[k,n], \quad (4)$$

one for each *time-frequency slot* (k,n). The coefficients $\{w[k,n]\}$ are independent and identically distributed (i.i.d.) JPG with zero mean and variance normalized to one.

C. Extension to Multiple Transmit and Receive Antennas

We extend the SISO channel model in (4) to a MIMO channel model with M_T transmit antennas, indexed by q, and M_R receive antennas, indexed by r, and assume that all component channels are characterized by the same scattering function $C_{\mathbb{H}}(\nu,\tau)$, so that they are diagonalized by the same set $\{g_{k,n}(t)\}$. For each slot (k,n) and component channel (r,q), the resulting scalar channel coefficient is denoted as $h_{r,q}[k,n]$. We allow for spatial correlation according to the *separable correlation* model [12]:

$$\mathbb{E}\left[h_{r,q}[k'+k,n'+n]h^*_{r',q'}[k',n']\right] = B[r,r']A[q,q']R[k,n].$$

The $M_T \times M_T$ matrix $\mathbf{A}$ with entries $[\mathbf{A}]_{q,q'} = A[q,q']$ is called the *transmit correlation matrix*, and the $M_R \times M_R$ matrix $\mathbf{B}$, with entries $[\mathbf{B}]_{r,r'} = B[r,r']$, is the *receive correlation matrix*. We denote by $\{\sigma_q\}$ and by $\{\lambda_r\}$ the set of eigenvalues of the matrices $\mathbf{A}$ and $\mathbf{B}$, respectively; the eigenvalues are ordered according to $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{M_T-1}$ and $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{M_R-1}$, and satisfy the normalization

$$\sum_{q=0}^{M_T-1} \sigma_q = M_T, \qquad \sum_{r=0}^{M_R-1} \lambda_r = M_R. \tag{5}$$

We define a *channel use* as a $K \times N$ rectangle of time-frequency slots and stack the symbols $\{x_q[k,n]\}$ transmitted from all M_T transmit antennas during one channel use into an $M_T KN$-dimensional vector $\mathbf{x}$, the corresponding output $\{y_r[k,n]\}$ for all M_R receive antennas into an $M_R KN$-dimensional vector $\mathbf{y}$, and likewise the noise $\{w_r[k,n]\}$ into an $M_R KN$-dimensional vector $\mathbf{w}$. Stacking proceeds first along frequency, then along time, and finally along space, as shown exemplarily for the input vector $\mathbf{x}$:

$$\mathbf{x}_q[k] = [x_q[k,0]\, x_q[k,1] \, \cdots \, x_q[k,N-1]]^T \tag{6a}$$

$$\mathbf{x}_q = [\mathbf{x}_q^T[0]\, \mathbf{x}_q^T[1] \, \cdots \, \mathbf{x}_q^T[K-1]]^T \tag{6b}$$

$$\mathbf{x} = [\mathbf{x}_0^T \, \mathbf{x}_1^T \, \cdots \, \mathbf{x}_{M_T-1}^T]^T. \tag{6c}$$

Analogously, we stack the channel coefficients, first in frequency to obtain the vectors $\mathbf{h}_{r,q}[k]$, and then in time to obtain a vector $\mathbf{h}_{r,q}$ for each component channel (r,q); further stacking of these vectors along transmit antennas q and then along receive antennas r results in the $M_T M_R KN$-dimensional vector $\mathbf{h}$. Let $\mathbf{X}_q = \operatorname{diag}\{\mathbf{x}_q\}$ and $\mathbf{X} = [\mathbf{X}_0 \, \mathbf{X}_1 \, \cdots \, \mathbf{X}_{M_T-1}]$, where the vectors $\mathbf{x}_q$ are defined in (6b). With this notation, the IO relation for one channel use can be conveniently expressed as

$$\mathbf{y} = (\mathbf{I}_{M_R} \otimes \mathbf{X})\mathbf{h} + \mathbf{w}. \tag{7}$$

The distribution of the channel coefficients in a given channel use is completely characterized by the $M_T M_R KN \times M_T M_R KN$ correlation matrix

$$\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] = \mathbf{B} \otimes \mathbf{A} \otimes \mathbf{R} \tag{8}$$

where the correlation matrix $\mathbf{R} = \mathbb{E}[\mathbf{h}_{r,q}\mathbf{h}_{r,q}^H]$ is the same for all component channels (r,q). Knowledge of the channel law means that the three matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{R}$ are known to the transmitter and the receiver.

We impose a constraint on the average power of the transmitted signal per channel use such that[1] $\mathbb{E}\left[\|\mathbf{x}\|^2\right]/T \leq KP$. In addition, we assume a peak constraint across transmit antennas in each slot (k,n) according to

$$\frac{1}{T}\sum_{q=0}^{M_T-1} |x_q[k,n]|^2 \leq \frac{\beta P}{N}, \quad \text{with probability 1.} \tag{9}$$

Here, $\beta \geq 1$ is the peak- to average-power ratio.

III. Capacity Bounds

With the system model and power constraints in place, we can now proceed to state our upper and lower bounds on the capacity of the channel with IO relation (7). As we assume that for all (r,q) the process $\{h_{r,q}[k,n]\}$ has a spectral density, given in (3), the channel process is ergodic in k for all component channels [16], and the capacity is given by

$$C(W) = \lim_{K\to\infty} \frac{1}{KT} \sup_{\mathcal{P}} I(\mathbf{y};\mathbf{x}) \tag{10}$$

for any fixed bandwidth $W = NF$. The supremum is taken over the set $\mathcal{P}$ of all input distributions that satisfy the constraints on peak and average power in Section II-C.

A. Upper Bound

Theorem 1: The capacity (10) of the underspread WSSUS MIMO channel in Section II-C, under the average-power constraint $\mathbb{E}\left[\|\mathbf{x}\|^2\right]/T \leq KP$ and the peak constraint (9), is upper-bounded as $C(W) \leq U_1(W)$, where

$$U_1(W) = \sup_{0 \leq \alpha \leq \sigma_0} \sum_{r=0}^{M_R-1} \left(\frac{W}{TF} \log\left(1 + \alpha\lambda_r \frac{PTF}{W}\right) - \alpha G_r(W) \right) \tag{11a}$$

$$G_r(W) = \frac{W}{\sigma_0 \beta} \iint_{\tau\,\nu} \log\left(1 + \frac{\sigma_0 \lambda_r \beta P}{W} C_{\mathbb{H}}(\nu,\tau)\right) d\nu d\tau. \tag{11b}$$

As the value of α that achieves the supremum in (11a) depends on W in general, the upper bound $U_1(W)$ is difficult to interpret. However, for the special case that the supremum is attained for $\alpha = \sigma_0$ independently of W, the upper bound can be interpreted as the capacity of a set of M_R parallel AWGN channels with received power $\sigma_0 \lambda_r P$ and $W/(TF)$ degrees of freedom per second, minus a penalty that quantifies the capacity loss due to channel uncertainty. It can be shown [14] that a sufficient condition for the supremum in (11a) to be achieved for $\alpha = \sigma_0$ is

$$\Delta_{\mathbb{H}} \leq \frac{\beta}{3TF} \quad \text{and} \quad \frac{P}{W} < \frac{\Delta_{\mathbb{H}}}{\sigma_0 \lambda_0 \beta}\left[\exp\left(\frac{\beta}{2TF\Delta_{\mathbb{H}}}\right) - 1\right].$$

As virtually all wireless channels are highly underspread, and as $\beta \geq 1$ and, typically, $TF \approx 1.25$ [15], the first condition above is always satisfied. Hence, only the second condition is relevant; but even for large spread $\Delta_{\mathbb{H}}$, this condition holds for all SNR values P/W of practical interest. As an example, consider a system with $\beta = 1$ and $M_T = M_R = 4$ antennas that operates over a channel with spread $\Delta_{\mathbb{H}} = 10^{-2}$. If we use the upper bound $\sigma_0 \lambda_0 \leq M_R M_T$, which follows from the normalization (5), we find that $P/W < 141\,\text{dB}$ is sufficient for the supremum in (11a) to be achieved for $\alpha = \sigma_0$. This value is far in excess of the receive SNR encountered in practical systems. Therefore, we exclusively consider the case $\alpha = \sigma_0$ in the remainder of the paper.

What we call the "penalty term", i.e., $\sigma_0 \sum_{r=0}^{M_R-1} G_r(W)$ in (11), is a lower bound on the mutual information between the channel $\mathbf{h}$ and the output $\mathbf{y}$, given the input $\mathbf{x}$ [14]. For SISO channels, it is shown in [15] that of all unit-volume scattering functions with prescribed ν_0 and τ_0, the brick-shaped scattering function, $C_{\mathbb{H}}(\nu,\tau) = 1/\Delta_{\mathbb{H}}$ for $(\nu,\tau) \in [-\nu_0,\nu_0] \times [-\tau_0,\tau_0]$, yields the largest penalty term. The same

[1] Current regulations for ultrawideband systems allow the average power to increase with bandwidth. As we keep the average power fixed irrespectively of the bandwidth, our results do not apply to such regulations.

is true for the MIMO channel at hand, and the corresponding capacity is upper-bounded as

$$C(W) \le \sum_{r=0}^{M_R-1} \left\{ \frac{W}{TF} \log\left(1 + \sigma_0 \lambda_r \frac{PTF}{W}\right) - \frac{W\Delta_{\mathbb{H}}}{\beta} \log\left(1 + \sigma_0 \lambda_r \frac{\beta P}{W\Delta_{\mathbb{H}}}\right)\right\}.$$

B. Lower Bound

Theorem 2: Let $\mathbf{C}(\theta)$ denote the $N \times N$ matrix-valued spectral density of an arbitrary component channel[2] $\{\mathbf{h}[k]\}$, i.e.,

$$\mathbf{C}(\theta) = \sum_{k=-\infty}^{\infty} \mathbb{E}\left[\mathbf{h}[k'+k]\mathbf{h}^H[k']\right] e^{-j2\pi k\theta}, \quad |\theta| \le \frac{1}{2}.$$

Furthermore, let $\mathbf{s}$ denote an M_T-dimensional vector whose first Q elements are i.i.d., have zero mean, and are of *constant modulus*, i.e., satisfy $|[\mathbf{s}]_q|^2 = PT/(QN)$, and let the remaining $M_T - Q$ elements be zero. Let $\mathbf{H}_w$ be an $M_R \times M_T$ matrix and $\mathbf{w}$ an M_R-dimensional vector, both with i.i.d. JPG entries of zero mean and unit variance. Finally, let $\mathbf{\Sigma} = \operatorname{diag}\{[\sigma_0\, \sigma_1\, \cdots\, \sigma_{M_T-1}]^T\}$ and $\mathbf{\Lambda} = \operatorname{diag}\{[\lambda_0\, \lambda_1\, \cdots\, \lambda_{M_R-1}]^T\}$. Denote by $I(\mathbf{s}; \mathbf{y}\,|\,\mathbf{H}_w)$ the coherent mutual information of the fading MIMO channel with IO relation $\mathbf{y} = \mathbf{\Lambda}^{1/2}\mathbf{H}_w\mathbf{\Sigma}^{1/2}\mathbf{s} + \mathbf{w}$. Then, the capacity (10) of the underspread WSSUS MIMO channel in Section II-C, under the average-power constraint $\mathbb{E}\left[\|\mathbf{x}\|^2\right]/T \le KP$ and the peak constraint (9), is lower-bounded as $C(W) \ge \max_{1\le Q\le M_T} L_1(W,Q)$, where

$$L_1(W,Q) = \max_{1\le\gamma\le\beta} \left\{ \frac{W}{\gamma TF} I(\mathbf{y}; \sqrt{\gamma}\,\mathbf{s}\,|\,\mathbf{H}_w) - \frac{1}{\gamma T} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \log\det\left(\mathbf{I}_N + \sigma_q \lambda_r \frac{\gamma PTF}{QW} \mathbf{C}(\theta)\right) d\theta \right\}.$$

For large enough bandwidth, and hence large enough N, the lower bound in Theorem 2 can be well approximated by an expression that is often much easier to evaluate: (i) we replace the first term of $L_1(W,Q)$ by its Taylor series up to first order, as given in [17, Theorem 3]; (ii) in the second term, we replace the $N \times N$ Toeplitz matrix $\mathbf{C}(\theta)$ by a circulant matrix that is asymptotically equivalent, in N, to $\mathbf{C}(\theta)$ [15]. The resulting wideband approximation of $L_1(W,Q)$ then reads

$$L_1(W,Q) \approx L_a(W,Q) = \max_{1\le\gamma\le\beta} \left\{ \frac{M_R P}{Q} \sum_{q=0}^{Q-1} \sigma_q - \gamma P^2 \frac{TF}{W} \frac{\left(\sum_{q=0}^{Q-1}\sigma_q\right)^2 \sum_{r=0}^{M_R-1}\lambda_r^2 + M_R^2 \sum_{q=0}^{Q-1}\sigma_q^2}{2Q^2} - \frac{W}{\gamma} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \iint_{\tau\,\nu} \log\left(1 + \sigma_q \lambda_r \frac{\gamma P}{QW} C_{\mathbb{H}}(\nu,\tau)\right) d\nu d\tau \right\}.$$

This approximation is exact for $W \to \infty$ [15].

[2]The vector processes $\{\mathbf{h}_{r,q}[k]\}$ of all component channels (r,q) have the same spectral density by assumption; therefore, we drop the subscripts r and q.

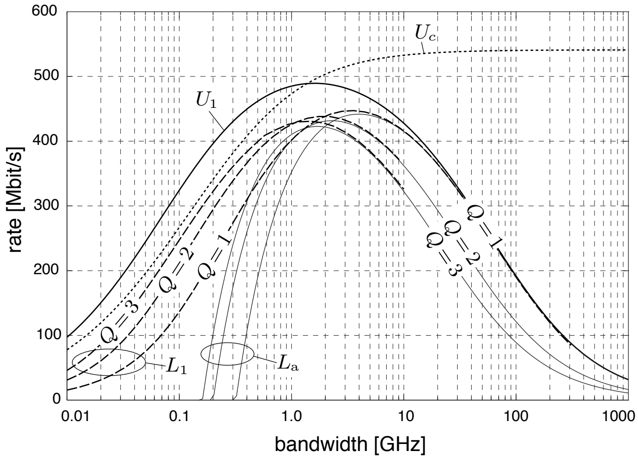

Fig. 1. Upper and lower bounds on the capacity of a spatially uncorrelated underspread WSSUS channel, $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\mathbb{H}} = 10^{-3}$.

C. Numerical Examples

For a 3×3 MIMO system, we show in Fig. 1 plots of the upper bound $U_1(W)$ in Theorem 1, and—for Q between 1 and 3—plots of the lower bound[3] $L_1(W,Q)$ in Theorem 2 and of the corresponding approximation $L_a(W,Q)$. For comparison, we also plot the standard coherent capacity upper bound $U_c(W)$ obtained for input subject to an average-power constraint only. We consider only the spatially uncorrelated case, i.e, $\mathbf{A} = \mathbf{B} = \mathbf{I}_3$; numerical results for spatially correlated channels can be found in [14].

All plots are for a receive power, normalized with respect to the noise spectral density, of $P/(1\,\mathrm{W/Hz}) = 1.26 \cdot 10^8 \mathrm{s}^{-1}$. This value corresponds, for example, to a transmit power of $0.5\,\mathrm{mW}$, a thermal noise level at the receiver of $-174\,\mathrm{dBm/Hz}$, free-space path loss over a distance of $10\,\mathrm{m}$, and a rather conservative receiver noise figure of $20\,\mathrm{dB}$. Furthermore, we assume that the scattering function is brick-shaped with $\tau_0 = 5\,\mu\mathrm{s}$, $\nu_0 = 50\,\mathrm{Hz}$, and corresponding spread $\Delta_{\mathbb{H}} = 10^{-3}$. Finally, we set $\beta = 1$.

We can observe that $U_c(W)$ is tighter than $U_1(W)$ for small bandwidth; this holds true in general in the spatially uncorrelated case, as for small W the penalty term in (11) can be neglected and we have that $U_1(W) \approx [M_R W/(TF)] \log(1 + PTF/W)$, which is the Jensen upper bound on the coherent capacity $U_c(W)$. For small and medium bandwidth, $L_1(W,Q)$ increases with Q and comes surprisingly close to $U_c(W)$ for $Q = 3$. As can be expected in the light of, e.g., [4], [5], when bandwidth increases above a certain *critical bandwidth*, $U_1(W)$ and $L_1(W,Q)$ start to decrease; in this regime, the rate gain obtained from the additional degrees of freedom is offset by the resources required to resolve channel uncertainty. The same argument seems to hold for spatial degrees of freedom: above a certain bandwidth, $U_1(W)$ appears to match $L_1(W,Q)$ for $Q = 1$; hence, using a single transmit antenna seems to be optimal in the wideband regime. We make this statement precise in the next section.

[3]Methods to numerically evaluate $L_1(W,Q)$ are discussed in [14], [15].

IV. The Wideband Regime

Fig. 1 suggests that in the wideband regime it is optimal to use a single transmit antenna when the channel is spatially uncorrelated both at the transmitter and the receiver side. To substantiate this observation and to understand the impact of transmit and receive correlation in the wideband regime, we compute the first-order Taylor series expansion of $C(W)$ around $1/W = 0$.

Theorem 3: Define

$$\kappa_{\mathbb{H}} = \iint_{\tau\,\nu} C^2_{\mathbb{H}}(\nu,\tau)d\nu d\tau, \qquad \text{and} \qquad \theta = \sum_{r=0}^{M_R-1} \lambda_r^2. \quad (12)$$

Then, for[4] $\beta > 2TF/\kappa_{\mathbb{H}}$, the capacity (10) of the underspread WSSUS MIMO channel in Section II-C, under the average-power constraint $\mathbb{E}\left[\|\mathbf{x}\|^2\right]/T \le KP$ and the peak constraint (9), has the following first-order Taylor series expansion around $1/W = 0$:

$$C(W) = \frac{a}{W} + o\left(\frac{1}{W}\right) \quad (13a)$$

where

$$a = \theta\,\frac{(\sigma_0 P)^2}{2}\left(\beta\kappa_{\mathbb{H}} - TF\right). \quad (13b)$$

The coefficient a of the first-order term of the Taylor series expansion of capacity in (13a) depends on the transmit correlation matrix $\mathbf{A}$, which we assume known at the transmitter, only through its maximum eigenvalue σ_0. We show in [14] that rank-one statistical beamforming along any of the eigenvectors of $\mathbf{A}$ associated with σ_0 is capacity optimal in the wideband regime. For channels that are spatially uncorrelated at the transmitter, this result implies that using only one transmit antenna is optimal, as previously shown in [6] for the time-selective, frequency-flat case.

To further assess the impact of spatial correlation on capacity, we follow [12], [18], [7], [9] and use standard tools from majorization theory [19]. In the *coherent setting*, capacity is Schur concave in the eigenvalue vector of the receive correlation matrix, while, for sufficiently large bandwidth, it is Schur convex in the eigenvalue vector of the transmit correlation matrix [18]. Hence, in the coherent setting, receive correlation is detrimental at any bandwidth, while transmit correlation is beneficial at large bandwidth [20].

On the basis of Theorem 3, we conclude that the picture is fundamentally different in the *noncoherent setting*. The coefficient a in (13b) is a Schur-convex function in both the eigenvalue vector $[\sigma_0\,\sigma_1\,\cdots\,\sigma_{M_T-1}]^T$ of the transmit correlation matrix and the eigenvalue vector $[\lambda_0\,\lambda_1\,\cdots\,\lambda_{M_R-1}]^T$ of the receive correlation matrix [14]. Hence, both transmit and receive correlation are beneficial for sufficiently large bandwidth. This observation agrees with the results for memoryless and block-fading channels reported in [7]–[10]. In the wideband regime, while transmit correlation is beneficial in both the coherent and the noncoherent setting because it allows for power focusing, receive correlation is beneficial rather than detrimental in the noncoherent setting for the following reason: for fixed M_T and M_R, the rate gain obtained from additional bandwidth is offset in the wideband regime by the corresponding increase in channel uncertainty (see Fig. 1); yet for fixed but large bandwidth, channel uncertainty decreases in the presence of receive correlation so that capacity increases.

[4] This condition holds for virtually all wireless channels of practical interest [14].

References

[1] U. G. Schuster and H. Bölcskei, "Ultrawideband channel modeling on the basis of information-theoretic criteria," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2464–2475, Jul. 2007.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, U.S.A.: Wiley, 1968.

[3] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400, Jul. 2000.

[4] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.

[5] V. G. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 809–827, Apr. 2002.

[6] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, 2008, submitted. [Online]. Available: http://arxiv.org/abs/0712.2872

[7] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[8] W. Zhang and J. N. Laneman, "Benefits of spatial correlation for multi-antenna non-coherent communication over fading channels at low SNR," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 887–896, Mar. 2007.

[9] S. G. Srinivasan and M. K. Varanasi, "Optimal spatial correlation for the noncoherent MIMO Rayleigh fading channel," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3760–3769, Oct. 2007.

[10] X. Wu and R. Srikant, "MIMO channels in the low-SNR regime: Communication rate, error exponent, and signal peakiness," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1290–1309, Apr. 2007.

[11] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[12] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[13] W. Kozek, "Matched Weyl-Heisenberg expansions of nonstationary environments," Ph.D. dissertation, Vienna University of Technology, Department of Electrical Engineering, Vienna, Austria, Mar. 1997.

[14] U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor, "Capacity bounds for peak-constrained multiantenna wideband channels," *IEEE Trans. Commun.*, Jan. 2008, submitted. [Online]. Available: http://arxiv.org/abs/0801.1002

[15] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, Apr. 2008, submitted. [Online]. Available: http://arxiv.org/abs/0804.1748

[16] G. Maruyama, "The harmonic analysis of stationary stochastic processes," *Memoirs of the Faculty of Science, Kyūshū University, Ser. A*, vol. 4, no. 1, pp. 45–106, 1949.

[17] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[18] E. A. Jorswieck and H. Boche, "Performance analysis of MIMO systems in spatially correlated fading using matrix-monotone functions," *IEICE Trans. Fund. Elec. Commun. Comp. Sc.*, vol. E89-A, no. 5, pp. 1454–1482, May 2006.

[19] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, U.S.A.: Academic Press, 1979.

[20] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.