On Non-coherent MIMO Capacity in the Wideband Regime: Interplay of SNR, Antennas and Coherence *

Siddharth Ray† Muriel Médard‡ Lizhong Zheng‡


Abstract

We consider a multiple-input, multiple-output (MIMO) wideband Rayleigh block fading channel, where the channel state is unknown to both the transmitter and the receiver. With only an average power constraint, we compute the capacity of this channel and consider its interaction with the coherence length, number of transmit and receive antennas and receive signal-to-noise ratio (SNR) per degree of freedom. We establish how large the coherence length has to be in order for the non-coherent channel to have a “near coherent” performance in the wideband regime. More specifically, we show that if the coherence length of the channel is above a certain SNR (bandwidth) dependent threshold, the non-coherent and coherent capacities are the same in the large bandwidth regime. We also propose a signaling scheme that is near-optimal in the wideband regime.

For the case when the channel state is perfectly known at the receiver, we analyze the effect of multiple antennas on the decoding error probability in the wideband limit. We show that though multiple transmit antennas do not improve the wideband capacity limit, they do reduce the error probability.


†S. Ray (sray@mit.edu), M. Médard (medard@mit.edu) and L. Zheng (lizhong@mit.edu) are with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139. Work is supported by NSF Ultra Wideband Wireless Award ANI0335256, NSF Career Award CCR 0093349 and Hewlett-Packard Award 008542-008.
1 Introduction

Recent years have seen the emergence of high data rate, third generation wideband wireless communication standards like wideband code division multiple access (W-CDMA) and Ultra-wideband (UWB) radio. Motivated by the ever increasing demand for higher wideband wireless data rates, we consider multiple antenna communication over the wideband wireless channel.

At the cost of additional signal processing (which is getting cheaper with rapid advances in VLSI technology), multiple-input multiple-output (MIMO) systems have been known to improve considerably performance of wireless systems in terms of reliability as well as throughput, without requiring additional resources such as bandwidth and power. However, multiple antenna research has focused primarily in the regime where the received signal-to-noise ratio (SNR) per degree of freedom is high. Such a regime operates in essence as a narrowband regime. We now study the performance of MIMO at the other extreme, i.e., when the available bandwidth is large, which takes us to the regime where the SNR per degree of freedom is low.

In wideband channels, the available power is spread over a large number of degrees of freedom. This makes the SNR per degree of freedom low. Hence, while studying these channels, we need to focus on the low SNR regime. We will therefore use the terms “wideband capacity” and “low SNR capacity” interchangeably, with the understanding that the latter refers to the capacity per degree of freedom.

In this paper, we assume that the transmitter has no channel state information (CSI). The receiver may or may not have CSI. When the receiver has perfect CSI, it is called a coherent channel, whereas, when it has no CSI, it is called a non-coherent channel. We also assume Rayleigh block fading. In the limit of infinite bandwidth, Zheng and Tse [12] show that the capacities per degree of freedom for the coherent and non-coherent MIMO channels are the same, i.e.,

$$\lim_{\text{SNR} \to 0} \frac{C_{\text{coherent}}(\text{SNR})}{\text{SNR}} = \lim_{\text{SNR} \to 0} \frac{C_{\text{non-coherent}}(\text{SNR})}{\text{SNR}} = r,$$

where, $r$ is the number of receive antennas and $\text{SNR}$ is the average signal-to-noise ratio per degree of freedom at each receive antenna. The capacity can thus be expressed as:

$$C(\text{SNR}) = r\text{SNR} + o(\text{SNR}) \text{ nats/channel use}$$

and is thus a linear function only in the limit of low SNR. As SNR increases from 0, capacity increases in a sublinear fashion, showing that low SNR communication is power efficient. Using a Taylor series expansion, Verdu [17] shows that the second derivative of the capacity at $\text{SNR} = 0$ is finite for the coherent channel. The impact on the coherent capacity of antenna
correlation, Ricean factors, polarization diversity and out-of-cell interference is considered in [22]. For the non-coherent channel, Verdú [17] shows that though “flash” signaling is first order optimal, it renders the second derivative $-\infty$. Hence, the coherent and non-coherent channels have the same linear term and differ in their sublinear term. Therefore, the non-coherent channel capacity approaches the wideband limit slower than the coherent channel capacity.

Let us define the sublinear term for the MIMO channel with $t$ transmit and $r$ receive antennas as

$$\Delta^{(t,r)} \text{SNR} = r \text{SNR} - C(\text{SNR}) \text{ nats/channel use.}$$

Computing the sublinear term tells us the capacity and also quantifies the convergence of the capacity function to the low SNR limit. Larger the order of the sublinear term, faster the convergence. Using the results of Verdú [17], the sublinear term for the Rayleigh fading coherent MIMO channel, $\Delta^{(t,r)}_{\text{coherent}}(\text{SNR})$, is

$$\Delta^{(t,r)}_{\text{coherent}}(\text{SNR}) = \frac{r(r + t)}{2t} \text{SNR}^2 + o(\text{SNR}^2).$$

On the other extreme, for the i.i.d Rayleigh fading non-coherent MIMO channel, the sublinear term, $\Delta^{(t,r)}_{\text{i.i.d}}(\text{SNR}) \gg O(\text{SNR}^2)$ [17]. In this paper, we compute the sublinear term for this
channel as

\[ \Delta_{i.i.d}^{(t,r)}(\text{SNR}) \triangleq \frac{r \cdot \text{SNR}}{\log(\frac{\text{SNR}}{r})}. \]

(The notation \( \triangleq \) denotes an approximation that ignores high order logarithm functions and is described in detail in Appendix 4.) We also show that on-off signaling achieves the capacity of this channel (with our approximation in mind).

Figure 1 shows the sublinear terms for the Rayleigh fading coherent channel and the i.i.d. Rayleigh fading non-coherent channel with the same number of transmit and receive antennas. A property of the non-coherent capacity is that it tends towards the coherent capacity as the coherence length increases. Hence, the sublinear term for the i.i.d. Rayleigh fading non-coherent channel is the largest (non-coherent extreme), whereas, for the coherent channel, it is the smallest (coherent extreme). In this paper, we focus on how the non-coherent MIMO channel capacity is influenced by the coherence length, number of antennas and SNR. We do so, by computing the sublinear term, which in turn tells us the capacity of the low SNR non-coherent MIMO channel of arbitrary coherence length. Thereby, we sweep the region, shown in Figure 1, between the coherent and non-coherent extremes.

In the low SNR regime, the sublinear term also represents the energy efficiency of communication. Let \( E_n \) and \( N_0 \) represent the energy per information nat and the noise spectral level, respectively. We have:

\[
\frac{E_n}{N_0} = \frac{\text{SNR}}{C(\text{SNR})} = \frac{\text{SNR}}{r \cdot \text{SNR} - \Delta_{i.i.d}^{(t,r)}(\text{SNR})} = \frac{1}{r} \cdot \frac{1}{1 - \frac{\Delta_{i.i.d}^{(t,r)}(\text{SNR})}{r \cdot \text{SNR}}}.
\]

Taking logarithms on both sides,

\[
\log \left( \frac{E_n}{N_0} \right) \approx \frac{\Delta_{i.i.d}^{(t,r)}(\text{SNR})}{r \cdot \text{SNR}} - \log(r). \tag{1}
\]

Equation (1) shows how energy efficiency is related to the sublinear term. The smaller the sublinear term for a channel, the more energy efficient will it be. As the non-coherent capacity is always less than the coherent capacity for the same number of transmit and receive antennas, lack of receiver CSI results in energy inefficiency. Also, note that the minimum energy (in dB) required to reliably transmit one information nat decreases logarithmically with the number of receive antennas.

Let us now turn to Figure 2, which shows how wideband capacity changes with bandwidth. We denote \( P \) as the average receive power and \( N_0 \) as the noise spectral density, which makes
the wideband limit \( \frac{rP}{N_0} \) nats/sec. We obtain this figure by scaling the y-axis of Figure 1 by the bandwidth. Channels whose capacities converge slowly to the wideband limit have to incur large bandwidth penalties. For the same number of transmit and receive antennas, the non-coherent capacity is less than the coherent capacity. Thus, the non-coherent channel requires a larger bandwidth in order to reliably support the same throughput as the coherent channel. This bandwidth penalty grows with bandwidth. Hence, for the non-coherent channel, the closer we get to the wideband limit, we gain in terms of energy efficiency (as the sublinear term decreases), but the bandwidth penalty becomes larger. We quantify this effect by computing the low SNR non-coherent MIMO capacity. Studying how capacity changes with coherence length also tells us the amount of bandwidth required to achieve a “near coherent” performance. Note that since bandwidth penalty increases with decreasing coherence length, the channel at the non-coherent extreme (i.i.d Rayleigh fading non-coherent channel) has to incur the largest bandwidth penalty.

At low SNR, channel estimates become unreliable. Hence, even for slowly varying channels, estimating the channel at the receiver may not be possible. When we have multiple antennas at the receiver as well as the transmitter, estimation becomes much more difficult since there are multiple channel coefficients that need to be estimated. Hence, communication is desirable without training. In [16], non-coherent communication is considered with the input distribution constrained to be exponentially decaying. It is show that the capacity per degree of freedom in the low SNR regime is \( O(\text{SNR}^2) \). Reference [21] considers the same capacity under the constraint that only the fourth and sixth order moments of the input are
finite. Once again, the non-coherent capacity per degree of freedom is shown to be $O(SNR^2)$. Hence, [16, 21] show that when there is a higher order (fourth and above) constraint on the input, capacity scales inversely with bandwidth. Thus, the non-coherent capacity does not approach the wideband limit and diverges from the coherent capacity as bandwidth increases. These results are akin to the single antenna channel results [10, 13, 14, 15]. Hassibi and Hochwald [20] propose a training scheme that increases the capacity beyond the non-coherent capacity in the high SNR regime. However, at low SNR, this scheme results in the rate per degree of freedom to go as $O(SNR^2)$. Since the overall rate decays to 0 inversely with bandwidth, this training scheme is not desirable at low SNR.

In this paper, we consider multiple antenna communication over a wideband, non-coherent Rayleigh block fading channel. We compute the capacity with only an average power constraint, and consider it’s interaction with the coherence length of the channel, number of transmit and receive antennas and SNR. We establish how large the coherence length has to be in order for the non-coherent channel to have a “near coherent” performance at low SNR. More specifically, we show that if the coherence length of the channel is above a certain SNR dependent threshold, the non-coherent and coherent capacities are the same in the low SNR regime. Moreover, we propose a signaling scheme that is near-optimal in the wideband regime.

The problem that we consider in this paper has been considered for single antenna channels by Zheng, Tse and Médard [25]. They consider the interaction between coherence length and capacity at low SNR and compute the order of the sublinear capacity term. They also propose a signaling scheme that is near-optimal in achieving this order. The work in this paper builds on their work, where, we analyze the more general MIMO channel and exactly compute the sublinear capacity term. The near-optimal signaling scheme that we propose is similar to the one obtained for the single antenna channel.

We also consider the effect of multiple antennas on the error probability of the coherent, low SNR MIMO channel. In the high SNR regime, the behavior of error probability is well understood for the coherent as well as the non-coherent channel [6, 9, 18, 19]. In the limit of infinite bandwidth, the behavior of the error exponent is analyzed in [23] for the non-coherent channel. In this work, we consider the case when the receiver has perfect CSI (coherent channel), and analyze the error probability in the wideband limit.

The paper is organized as follows. Section 2 discusses background and we describe our notation in Section 3. In Section 4, we describe our channel model and analyze the probability of error for the coherent MIMO channel in Section 5. In section 6, we compute the non-coherent MIMO channel capacity in the low SNR regime. We conclude in Section 7.
2 Background

The study of single antenna wideband channels dates back to 1969 and early work has considered the Rayleigh fading channel model. Kennedy [1] shows that the capacity of an infinite bandwidth Rayleigh fading channel is the same as that of an infinite bandwidth additive white Gaussian noise (AWGN) channel with the same average received power. Using the results of Gallager [3], Telatar [4] obtains the capacity per unit energy for the Rayleigh fading channel as a function of bandwidth and signal energy, concluding that given an average power constraint, the Rayleigh fading and AWGN channels have the same capacity in the limit of infinite bandwidth. Telatar and Tse [10] show that this property of the channel capacity is also found in channels with general fading distributions.

Médard and Gallager [13, 14] establish that very large bandwidths yield poor performance for systems that spread the available power uniformly over time and frequency (for example DS-CDMA). They express the input process as an orthonormal expansion of basis functions localized in time and frequency. The energy and fourth moment of the coefficients scale inversely with the bandwidth and square of the bandwidth, respectively. By constraining the fourth moment (as is the case when using spread spectrum signals), they show that mutual information decays to 0 inversely with increasing bandwidth. Telatar and Tse [10] consider a wideband fading channel to be composed of a number of time-varying paths and show that the input signals needed to achieve capacity must be “peaky” in time or frequency. They also show that if white-like signals are used (as for example in spread spectrum communication), the mutual information is inversely proportional to the number of resolvable paths with energy spread out and approaches 0 as the number of paths get large. This does not depend on whether the paths are tracked perfectly at the receiver or not. A strong coding theorem is obtained for this channel in [24]. Subramanian and Hajek [15] derive similar results as [13, 14] using the theory of capacity per unit cost, for a certain fourth-order cost function, called fourthergy.

We now consider the use of multiple antennas over these channels. MIMO channels were first studied from a capacity point of view in [5, 8]. In a Rayleigh flat-fading environment with perfect channel state information (CSI) at the receiver (coherent channel) but no CSI at the transmitter, and statistically independent propagation coefficients between all pairs of transmit and receive antennas, the multiple antenna capacity increases linearly with the smaller of the number of transmit and receive antennas, provided the signal-to-noise ratio is high [5].

When the coherence time of the channel is small (for example if the receiver is mobile), communication is desirable without training. Here, CSI is unavailable at the transmitter as well as the receiver. This channel is also referred to as the non-coherent channel. In [7], Marzetta and Hochwald derive the structure of the optimal input distribution as a product
of two statistically independent matrices; one of them being isotropically distributed and
the other being a diagonal, real and non-negative matrix. They also show that there is no
gain, from the point of view of capacity, in having the number of transmit antennas be more
than the coherence interval (in symbols) of the channel. Zheng and Tse [12] obtain the
non-coherent MIMO capacity in the high SNR regime and show that, in this regime, the
number of transmit antennas required need not be more than half the coherence interval (in
symbols).

3 Notation

In this section, we establish the notation we will be using throughout the paper. We will
use the bold type to denote random quantities whereas normal type for deterministic ones.
Matrices will be denoted by capital letters and the scalar or vector components of matrices
will be denoted using appropriate subscripts. Vectors will be represented by small letters
with an arrow over them. All vectors are column vectors unless they have a $^T$ superscript,
in which case they are row vectors. Scalars will be represented by small letters only. $^T$ will
be used to denote the complex conjugate transpose.

4 Model

We model the wideband channel as a set of $N$ parallel narrowband channels. We assume that
the Doppler spread is negligible, which makes the narrowband channels have independent
and identical statistics. Moreover, we assume that the coherence bandwidth is much larger
than the bandwidth of the narrowband channel. Hence, each narrowband channel is modeled
as being flat faded. Using the sampling theorem, the $m^{th}$ narrowband channel at symbol
time $k$ can be represented as:

$$\tilde{y}[k, m] = H[k, m]\tilde{x}[k, m] + \tilde{w}[k, m],$$

where $H[k, m]$, $\tilde{x}[k, m]$, $\tilde{w}[k, m]$ and $\tilde{y}[k, m]$ are the channel matrix, input vector, noise
vector and output vector, respectively, for the $m^{th}$ narrowband channel at symbol time $k$.
The pair $(k, m)$ may be considered as an index for the time-frequency slot, or degree of
freedom, to communicate. We denote the number of transmit and receive antennas by $t$
and $r$, respectively. Hence, $\tilde{x}[k, m] \in \mathbb{C}^t$ and $\tilde{y}[k, m], \tilde{w}[k, m] \in \mathbb{C}^r$. The channel matrix
$H[k, m]$ is a $r \times t$ complex matrix. The entries of the channel matrix are i.i.d, zero-mean,
complex Gaussian, with independent real and imaginary components. Equivalently, each
entry of $H[k, m]$ has uniformly distributed phase and Rayleigh distributed magnitude. We
thus model a Rayleigh fading channel with enough separation within the transmitting and
receiving antennas to achieve independence in the entries of $H[k, m]$. The channel matrix is
unknown at the transmitter but may be known at the receiver. However, its statistics are
known to both. The noise vector $\tilde{w}[k, m]$ is a zero-mean Gaussian vector with the identity
as its covariance matrix. Thus, $\tilde{w}[k, m] \sim \mathcal{CN}(0, I_r)$. Since the Doppler spread is negligible,
the narrowband channels are independent and we will omit the narrowband channel index, $m$, for simplifying notation. The capacity of the wideband channel with power constraint $P$ is thus $N$ times the capacity of each narrowband channel with power constraint $P/N$. Hence, we can focus on the narrowband channel alone.

We further assume a block fading channel model. Hence, the channel matrix is random but
fixed for the duration of the coherence time of the channel, and is i.i.d across blocks. Hence,
we may omit the time index, $k$, and express the narrowband channel within the $d^{th}$ coherence
block of length $l$ symbols as:

$$Y^d = H^d X^d + W^d, \quad d = 1, 2, \ldots$$

where, $X^d \in C^{t \times l}$ has entries $x_{ij}^d, i = 1, \ldots, t, j = 1, \ldots, l$, being the signals transmitted from
the transmit antenna $i$ at time $j$; $Y^d \in C^{r \times l}$ has entries $y_{ij}^d, i = 1, \ldots, r, j = 1, \ldots, l$, being the
signals received at the receive antenna $i$ at time $j$; the additive noise $W^d$ has i.i.d. entries
$w_{ij}^d$, which are distributed as $\mathcal{CN}(0, 1)$. The input $X$ satisfies the average power constraint

$$\frac{1}{l} E[\text{trace}(XX^\dagger)] = \text{SNR},$$

where, $\text{SNR}$ is the average signal to noise ratio at each receive antenna per narrowband
channel. As $N$ tends to $\infty$, $\text{SNR}$ tends to 0, and the narrow band channel is in the low SNR
regime.

5 Error Probability for the coherent MIMO channel

We now focus on the MIMO coherent channel at low SNR and our aim is to understand
the effect of multiple antennas on the error probability. We consider random coding over
$N$ successive blocks and denote the probability of decoding error as $P_e^N(R)$, where $R$ is the
transmission rate in nats/channel use. A decoding error is caused if either the channel matrix
is atypically ill-conditioned (type 1), which leads to outage, noise is atypically large (type
2), or some codewords are atypically close to each other (type 3). Type 1 errors, which lead
to outage, can also be described as the event where the mutual information of the channel
does not support a target data rate. Thus, the error probability can be upper bounded as

$$P_e^N(R) = P_{\text{error|outage}}^N(R) P_{\text{outage}}^N(R) + P_{\text{error|no outage}}^N(R)[1 - P_{\text{outage}}^N(R)]$$

$$\leq P_{\text{outage}}^N(R) + P_{\text{error|no outage}}^N(R).$$

(2)
We compute $P_{\text{outage}}^N(R)$ and $P_{\text{error|no outage}}^N(R)$.

5.1 Outage Probability

Let $X^d$ and $Y^d$ be the transmitted and received signals in the $d^{th}$ block. As the transmission rate is $R$ nats/channel use, the outage probability over $N$ blocks is

$$P_{\text{outage}}^N(R) = \inf_{q(X)} \Pr \left( \sum_{d=1}^{N} I(X^d; Y^d) < lNR \right),$$

where the minimization is over all possible distributions, $q(X)$, on $X$. Let us assume that the columns of $X$ are i.i.d zero mean complex Gaussian random vectors with covariance $Q = \frac{\text{SNR}}{t} I_t$. Though this distribution minimizes outage probability at high SNR [12], it does not do so in general [8]. However, by assuming a Gaussian input distribution, the analysis is simplified and we obtain an upper bound to the error probability. Equation (3) becomes

$$P_{\text{outage}}^N(R) \leq \Pr \left( \sum_{d=1}^{N} \log \det \left( I_r + \frac{\text{SNR}}{t} H^d H^d \right) < NR \right)$$

$$= \Pr \left( \sum_{d=1}^{N} \log \det \left( I_t + \frac{\text{SNR}}{t} H^d H^d \right) < NR \right)$$

$$= \Pr \left( \sum_{d=1}^{N} \log \prod_{s=1}^{m} \left( 1 + \frac{\text{SNR}}{t} \Lambda_s^d \right) < NR \right)$$

$$= \Pr \left( \sum_{d=1}^{N} \sum_{s=1}^{m} \log \left( 1 + \frac{\text{SNR}}{t} \Lambda_s^d \right) < NR \right)$$

$$\approx \Pr \left( \sum_{d=1}^{N} \sum_{s=1}^{m} \frac{\text{SNR}}{t} \Lambda_s^d < NR \right)$$

$$= \Pr \left( \frac{\text{SNR}}{t} \sum_{d=1}^{N} \text{trace}(H^d H^d) < NR \right).$$

In (4), $\Lambda_s^d$ is the $s^{th}$ unordered positive eigenvalue of the matrix $H^d H^d$ where $s \in \{1, ..., m\}$ and $m = \min(t, r)$. The approximation in (5) is valid since, we are considering low SNR. Define random variable $V$ as

$$V = \sum_{d=1}^{N} \text{trace}(H^d H^d).$$
\( V \) is an \textit{Erlang} distribution of order \( trN \). The cumulative distribution function of \( V \) can be written as

\[
F_V(v) = 1 - \left[ \sum_{i=0}^{trN-1} \frac{v^i}{i!} \right] \exp(-v).
\]

Thus, (6) becomes

\[
P_{\text{outage}}^N(R) \\
\leq \Pr \left( V \leq \frac{tNR}{\text{SNR}} \right) \\
= F_V \left( \frac{tNR}{\text{SNR}} \right) \\
= 1 - \left[ \sum_{i=0}^{trN-1} \frac{1}{i!} \left( \frac{tNR}{\text{SNR}} \right)^i \right] \exp \left( - \frac{tNR}{\text{SNR}} \right) \\
= \left[ \sum_{i=trN}^{\infty} \frac{1}{i!} \left( \frac{tNR}{\text{SNR}} \right)^i \right] \exp \left( - \frac{tNR}{\text{SNR}} \right).
\]

Therefore, the outage probability at low SNR is upper bounded as

\[
P_{\text{outage}}^N(R) \leq \left[ \sum_{i=trN}^{\infty} \frac{1}{i!} \left( \frac{tNR}{\text{SNR}} \right)^i \right] \exp \left( - \frac{tNR}{\text{SNR}} \right). \tag{7}
\]

Let us transmit at a rate which is a fixed fraction, \( \mu \in (0, 1) \), of the low SNR ergodic capacity:

\[ R = \mu r \text{SNR}. \]

This makes

\[
P_{\text{outage}}^N(\mu r \text{SNR}) \leq \left[ \sum_{i=trN}^{\infty} \frac{(\mu trN)^i}{i!} \right] \exp(-\mu trN). \tag{8}
\]

Define

\[
f(\mu, t, r, N) \triangleq \left[ \sum_{i=trN}^{\infty} \frac{(\mu trN)^i}{i!} \right] \exp(-\mu trN). \tag{9}
\]

We establish the following lemma:

\textbf{Lemma 1} For \( \mu \in (0, 1) \),

\[
\frac{\partial f(\mu, t, r, N)}{\partial \mu} > 0, \tag{10}
\]

\[
f(\mu, t, r, N) \leq \frac{\mu}{(1 - \mu)^2} \left[ \frac{1}{trN} \right]. \tag{11}
\]
Proof: Differentiating $f(\mu, t, r, N)$ with respect to $\mu$, we obtain
\[
\frac{\partial f(\mu, t, r, N)}{\partial \mu} = \frac{(\mu trN)^{trN}}{\mu(trN - 1)!} \exp(-\mu trN).
\]
Since,
\[
\frac{(\mu trN)^{trN}}{\mu(trN - 1)!} \exp(-\mu trN) > 0 \quad \forall \mu \in (0, 1),
\] (12)
equation (10) is proved.
We now prove (11). Let $z_{trN}$ be a Poisson random variable with parameter $\mu trN$. We can express $f(\mu, t, r, N)$ as
\[
f(\mu, t, r, N) = \Pr(z_{trN} \geq trN).
\] (13)
The mean ($m_{z_{trN}}$) and variance ($\lambda_{z_{trN}}$) of $z_{trN}$ are
\[
m_{z_{trN}} = \mu trN,
\]
\[
\lambda_{z_{trN}} = \mu trN.
\]
Now,
\[
\Pr(z_{trN} \geq trN)
\]
Chebyshev’s inequality is used to obtain (14). Combining (13, 14), we have

\[
f(\mu, t, r, N) \leq \frac{\mu}{(1 - \mu)^2} \left[ \frac{1}{trN} \right].
\]

The proof of the lemma is now complete.

From (8,9) and Lemma 1, we can infer the following: The outage probability increases with \( \mu \), for \( \mu \in (0, 1) \). Thus, for fixed \( t, r \) and \( N \), the closer the transmission rate is to the ergodic capacity, the higher the outage probability. More importantly, we see that for fixed \( \mu \), the outage probability decreases as the number of transmit antennas increases. Thus, though multiple transmit antennas do not improve the wideband capacity limit, they do decrease the outage probability. Note that increasing the number of receive antennas, \( r \), and increasing \( N \) also decreases the outage probability. Setting \( r = 5 \) and \( N = 5 \), we obtain plots (Figure 3) of the outage probability \( P^N_{\text{outage}}(\mu r \text{SNR}) \) vs \( \mu \) for different values of \( t \).

### 5.2 Computing \( P^N_{\text{(error|no outage)}}(R) \)

We now compute the expression for \( P^N_{\text{(error|no outage)}}(R) \). We can think of the input matrix transmitted in a block, \( X \), as a super symbol of dimension \( t \times l \). In this way, the channel is memoryless, since for each use of the channel an independent realization of \( H \) is drawn. As the code spans \( N \) blocks, we have from the strong coding theorem [2],

\[
P^N_{\text{(error|no outage)}}(R) = \exp(-NE(R)),
\]

where the error exponent for one block, \( E(R) \), is given by

\[
E(R) = \max_{q(X), \sigma \in [0,1]} \left[ E_0(\sigma, q(X)) - \sigma Rl \right].
\]

\( q(X) \) is the distribution of the random input matrix \( X \) and \( R \) is the transmission rate in nats/channel use. Since our error exponent corresponds to the probability of error when there is no outage, \( E_0(\sigma, q_X) \) is given by

\[
E_0(\sigma, q(X)) = -\log \int \int \left[ \int q(X)p(Y, H|X) \frac{1}{trN} dX \right]^{1+\sigma} dYdH.
\]

13
Since, $H$ is independent of $X$,

$$p(Y, H|X) = p(H)p(Y|X, H),$$

and

$$E_0(\sigma, q(X)) = -\log \left( E_H \left[ \int \left[ \int q(X)p(Y|X, H) \frac{1}{1+\sigma} dX \right]^{1+\sigma} dY \right] \right).$$

Let us assume that the columns of $X$ are i.i.d zero mean complex Gaussian random vectors with covariance $Q = \frac{SNR}{t} I_t$. This distribution may not maximize the error exponent [8], but simplifies analysis. With this assumption, we obtain a lower error exponent and an upper bound on the probability of error. $E_0(\sigma, q_X)$ is now a function function only of $\sigma$ and we will represent it as $E_0(\sigma)$. Therefore

$$\max_{\sigma \in [0,1]} [E_0(\sigma) - \sigma R_l] \leq E(R).$$

We now compute $E_0(\sigma)$. The computation yields the following lemma:

**Lemma 2**

$$E_0(\sigma) = \frac{\sigma r_{SNR}}{1 + \sigma}$$

**Proof:** See Appendix 1. \hfill \square

Thus,

$$\max_{\sigma \in [0,1]} \left[ E_0(\sigma) - \sigma R_l \right] = \max_{\sigma \in [0,1]} \left[ \frac{\sigma r_{SNR}}{1 + \sigma} - \sigma R_l \right]$$

$$= \begin{cases} 
\frac{l(r_{SNR}^2 - R)}{2} & 0 \leq R \leq \frac{r_{SNR}}{4} \\
 l(\sqrt{r_{SNR}} - \sqrt{R})^2 & \frac{r_{SNR}}{4} \leq R \leq r_{SNR}. 
\end{cases}$$

For $0 \leq R \leq \frac{r_{SNR}}{4}$, the error exponent is maximized at $\sigma = 1$. Thus, using (17)

$$E(R) \geq \begin{cases} 
\frac{l(r_{SNR}^2 - R)}{2} & 0 \leq R \leq \frac{r_{SNR}}{4} \\
 l(\sqrt{r_{SNR}} - \sqrt{R})^2 & \frac{r_{SNR}}{4} \leq R \leq r_{SNR}.
\end{cases}$$

(18)
Note that the error exponent does not depend on $t$. Hence, increasing the number of transmit antennas does not change the error exponent for the channel with a well-conditioned channel matrix, and thus, has no impact on the errors due to the noise's being atypically large or some codewords' being atypically close to each other. However, the error exponent increases with coherence length, $l$. Thus, we have a lower error probability as the coherence length of the channel increases. We see that for any rate below capacity, as number of receive antennas increases, the error exponent increases and thus, the probability of error occurring due to atypically large noise or close codewords decreases. Figure 4 shows the plot of the error exponent for different values of $r$ with SNR at 0 dB and $l = 1$.

Combining (15) and (18), we obtain

\[
P^N_{\text{error(no outage)}}(R) 
\leq \begin{cases} 
\exp \left[ -Nl\left( \frac{r\text{SNR}}{2} - R \right) \right] & 0 \leq R \leq \frac{r\text{SNR}}{4} \\
\exp \left[ -Nl\left( \sqrt{r\text{SNR}} - \sqrt{R} \right)^2 \right] & \frac{r\text{SNR}}{4} \leq R \leq r\text{SNR}.
\end{cases}
\tag{19}
\]

Combining (2), (7) and (19), the upper bound for the error probability for $0 \leq R \leq \frac{r\text{SNR}}{4}$ is

\[
P^N_e(R) \leq \sum_{i=trN}^{\infty} \frac{1}{i!} \left( \frac{tNR}{\text{SNR}} \right)^i \exp \left( - \frac{tNR}{\text{SNR}} \right) + \exp \left[ -Nl\left( \frac{r\text{SNR}}{2} - R \right) \right],
\]

Figure 4: Plot of Error Exponent vs Rate for different values of $r$. 
and for \( \frac{r_{SNR}}{4} \leq R \leq r_{SNR} \) is

\[
P_c^N(R) \leq \left[ \sum_{i=t_{SNR}}^{\infty} \frac{1}{i!} \left( \frac{tNR}{SNR} \right)^i \right] \exp \left( - \frac{tNR}{SNR} \right) + \exp \left[ - Nl(\sqrt{r_{SNR}} - \sqrt{R})^2 \right].
\]

Therefore, we see that increasing the number of transmit antennas decreases the outage probability component of the error probability or in other words, reduces the chance of the channel matrix being atypically ill-conditioned. As we increase the coherence length, we reduce the probability of error due to atypically large noise and atypically close codewords. An increase in \( N \) or number of receive antennas, \( r \), reduces errors that can arise out of either the noise’s being atypically large, codewords’ being atypically close or the channel matrix being ill-conditioned.

6 Non-coherent MIMO channel

In this section, we compute the capacity of the non-coherent MIMO channel at low SNR. The analysis shows the interaction between the number of receive and transmit antennas, coherence length of the channel and SNR in the wideband regime. We also propose a peaky signaling scheme that achieves capacity.

6.1 Dependence of capacity on coherence length

We first analyze the dependence of the non-coherent capacity on the coherence length of the channel.

6.1.1 Structure of optimal input distribution

In [7], the structure of the capacity achieving input distribution for our non-coherent MIMO channel model is described as

\[
X = A\Phi,
\]

where

\[
A = \begin{bmatrix}
|\mathbf{x}_1^T| \\
|\mathbf{x}_2^T| \\
|\mathbf{x}_i^T| \\
|\mathbf{x}_N^T|
\end{bmatrix},
\]

(continued)
is a \( t \times l \) random matrix that is diagonal, real and nonnegative with identically (though may not be independent) distributed entries and \( \|\mathbf{x}_i^T\| \) is the norm of the signal vector transmitted by the \( i^{th} \) antenna. Since these entries are identically distributed, we have \( \forall i \in \{1, \ldots, t\} \)

\[
E[\|\mathbf{x}_i^T\|^2] = \frac{1}{l} \text{SNR}.
\]

\( \Phi \) is a \( l \times l \) isotropically distributed unitary matrix. The row vectors of \( \Phi \) are isotropic random vectors which represent the direction of the signal transmitted from the antennas. \( \mathbf{A} \) and \( \Phi \) are statistically independent matrices. Since this structure of the input distribution is optimal, we will restrict our attention to input distributions having such structure.

We first prove Lemma 3, which establishes two necessary conditions the input distribution must satisfy for the mutual information of the channel to be above a certain value. This lemma will be used later on to establish the dependence of the non-coherent capacity on the channel coherence length.

**Lemma 3** For any \( \alpha \in (0, 1] \) and \( \gamma \in (0, \alpha) \), if there exists an input distribution on \( \mathbf{X} \) such that

\[
\frac{1}{l} I(\mathbf{X}; \mathbf{Y}) \geq r \text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}),
\]

then the following two conditions are satisfied by this distribution:

\[
\frac{t}{l} E[\log(1 + \|\mathbf{x}_i^T\|^2)] \leq \frac{(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}),
\]

\[
\frac{t}{l} E\left[\log\left(1 + \frac{\|\mathbf{x}_i^T\|^2}{l}\right)\right] \geq \text{SNR} - \frac{(r + t)}{t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}),
\]

for all \( i \in \{1, \ldots, t\} \).

**Proof:** See Appendix 2. \( \square \)

We are now ready to prove a theorem that establishes the dependence of the non-coherent capacity on the coherence length:

**Theorem 1** Consider a non-coherent Rayleigh block fading channel with average signal to noise ratio \( \text{SNR} \). Let the block length be \( l \) and the capacity \( C(\text{SNR}) \). For any \( \alpha \in (0, 1] \) and \( \gamma \in (0, \alpha) \), if

\[
C(\text{SNR}) \geq C^*(\text{SNR}) \triangleq r \text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}),
\]

then

\[
l > l_{\text{min}} \triangleq \frac{l^2}{(r + t)^2} \text{SNR}^{-2\alpha}.
\]

17
This theorem states that the coherence length of the channel must be strictly above $l_{\min}$ in order for the capacity to be above $C^*(\text{SNR})$. Since the inequality for the coherence length is strict, this implies that a channel with capacity $C^*(\text{SNR})$ will have its coherence length, $l^*$, strictly greater than $l_{\min}$, i.e.

$$l_{\min} < l^*.$$  

### 6.2 Communicating using Gaussian-like signals

In this subsection, we propose a signalling scheme using which a rate of $C^*(\text{SNR})$ is achievable if the coherence length is greater than or equal to a threshold, which we denote as $l^G$.

We first prove a lemma that shows that using a Gaussian input distribution, we can achieve “near coherent” performance if the coherence length of the channel is large enough.

**Lemma 4** Consider a non-coherent Rayleigh block fading channel with average signal-to-noise ratio $\text{SNR}$. Let the block length be $l$ and the capacity $C(\text{SNR})$. If we use Gaussian signals over this channel, then for any $\epsilon \in (0, 1)$, if

$$l \geq \frac{l^2}{(r+t)^2}\text{SNR}^{-2(1+\epsilon)},$$

then

$$C(\text{SNR}) \geq r\text{SNR} - \frac{r(r+t)}{2t}\text{SNR}^2 + O(\text{SNR}^{2+\epsilon}).$$

**Proof:** We first lower bound the mutual information of the non-coherent channel MIMO channel as

$$I(X; Y) = I(X; Y|H) + I(H; Y) - I(H; Y|X) \geq I(X; Y|H) - I(H; Y|X). \quad (22)$$

Let us choose the distribution of $X$ to be one where all the entries of $X$ are i.i.d. and $\mathcal{CN}(0, \frac{\text{SNR}}{l})$. Note that it is exactly this distribution that achieves capacity for the coherent MIMO channel. Therefore,

$$\frac{1}{l}I(X; Y|H) = r\text{SNR} - \frac{r(r+t)}{2t}\text{SNR}^2 + O(\text{SNR}^3). \quad (23)$$
$I(H;Y|X)$ is the information that can be obtained about $H$ from observing $Y$, conditioned on $X$ being known. Therefore

$$I(H;Y|X) = h(Y|X) - h(Y|X,H)$$

$$= rtE \log(1 + \|x_f^T\|^2)$$

$$\leq rt \log \left(1 + \frac{l}{t} \text{SNR}\right), \quad (24)$$

where we have used Jensen’s inequality to get the upper bound in (24). Combining (22), (23) and (24) and noting that

$$C(\text{SNR}) \geq \frac{1}{l} I(X;Y),$$

we obtain:

$$C(\text{SNR}) \geq r \text{SNR} - \frac{r(r + t)^2}{2t} \text{SNR}^2 - \frac{r^2}{l} \log \left(1 + \frac{l}{t} \text{SNR}\right) + O(\text{SNR}^3). \quad (25)$$

For any $\epsilon \in (0,1]$, let us choose

$$l = \frac{t^2}{(r + t)^2} \text{SNR}^{-2(1+\epsilon)}.$$

Therefore,

$$r \frac{t}{l} \log \left(1 + \frac{l}{t} \text{SNR}\right)$$

$$= r \frac{(r + t)^2}{l} \text{SNR}^{2(1+\epsilon)} \log \left(1 + \frac{t}{(r + t)^2} \text{SNR}^{1+2\epsilon}\right)$$

$$= r \frac{(r + t)^2}{l} \text{SNR}^{2(1+\epsilon)} \log \left(\frac{t}{(r + t)^2} \text{SNR}^{1+2\epsilon}\right) + o(\text{SNR}^{2(1+\epsilon)})$$

$$= r \frac{(r + t)^2}{l} \text{SNR}^{2(1+\epsilon)} \log \left(\frac{t}{(r + t)^2}\right) + r \frac{(r + t)^2}{l} \left(1 + 2\epsilon\right) \text{SNR}^{2+\epsilon} \left[\text{SNR}^\epsilon \log \left(\frac{1}{\text{SNR}}\right)\right] + o(\text{SNR}^{2(1+\epsilon)})$$

$$\leq r \frac{(r + t)^2}{l} \text{SNR}^{2(1+\epsilon)} \log \left(\frac{t}{(r + t)^2}\right) + r \frac{(r + t)^2}{l} \left(1 + 2\epsilon\right) \text{SNR}^{2+\epsilon} + o(\text{SNR}^{2(1+\epsilon)}) \quad (26)$$

In (26), we use that, since $\epsilon > 0$ and $\text{SNR} \to 0$, $\text{SNR}^\epsilon \log(\frac{1}{\text{SNR}}) \ll 1$. Since $r \frac{t}{l} \log \left(1 + \frac{l}{t} \text{SNR}\right)$ decreases monotonically with $l$, we have that

$$l \geq \frac{t^2}{(r + t)^2} \text{SNR}^{-2(1+\epsilon)}$$

$$\Rightarrow r \frac{t}{l} \log \left(1 + \frac{l}{t} \text{SNR}\right) \leq O(\text{SNR}^{2+\epsilon}).$$
Combining this with (25) completes the proof.

We now introduce an input distribution that has a flashy, as well as a continuous, nature. Such type of input distributions were first introduced in [25] for achieving the order of the sublinear capacity term for a single-input, single-output, non-coherent Rayleigh block fading channel.

For a given \( \alpha \in (0, 1] \), let us transmit in only \( \delta(\text{SNR}) = \text{SNR}^{1-\alpha} \) fraction of the blocks. As we are in the low signal to noise ratio regime, \( \delta(\text{SNR}) \in (\text{SNR}, 1] \). Since we concentrate the power only over a fraction of the blocks, the signal to noise ratio for the blocks in which we transmit increases to \( \text{SNR}' \) where

\[
\text{SNR}' = \frac{\text{SNR}}{\delta(\text{SNR})} = \text{SNR}^\alpha.
\]

In the blocks that we choose to transmit, let the entries of the input matrix \( X \) be i.i.d. \( \mathcal{CN}(0, \frac{\text{SNR}}{t}) \).

Note that as we increase \( \alpha \) from 0 to 1, the fraction of blocks that we transmit increases from \( \text{SNR} \) to 1. Therefore, as \( \alpha \) increases, the distribution changes from a peaky to a continuous one. We will call this type of signalling as peaky Gaussian and will prove a theorem that uses this to communicate over the non-coherent MIMO channel.

**Theorem 2** Consider a non-coherent Rayleigh block fading channel with average signal to noise ratio \( \text{SNR} \). Let the block length be \( l \) and the capacity \( C(\text{SNR}) \). If we use peaky Gaussian signals over this channel, then for any \( \alpha \in (0, 1] \) and \( \epsilon \in (0, \alpha) \), if

\[
l \geq l^G = \frac{t^2}{(r + t)^2} \text{SNR}^{-2(\alpha+\epsilon)},
\]

then

\[
C(\text{SNR}) \geq C^*(\text{SNR}) = r\text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\epsilon}).
\]

**Proof:** Let us use the peaky Gaussian like distribution for communicating over the non-coherent MIMO channel. We can now apply Lemma 4 to the blocks that we choose to transmit. Note that these blocks have a signal to noise ratio of \( \text{SNR}' \). Thus, for any \( \epsilon' \in (0, 1] \), if

\[
l \geq \frac{t^2}{(r + t)^2} \text{SNR'}^{-2(1+\epsilon')}
\]

then

\[
C(\text{SNR}') \geq r\text{SNR'} - \frac{r(r + t)}{2t} \text{SNR'}^{2} + O(\text{SNR'}^{2+\epsilon'}).\]
Since we are transmitting in $\delta(\text{SNR})$ fraction of the blocks,

$$C(\text{SNR}) = \delta(\text{SNR})C(\text{SNR}')$$

$$\geq r \text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\epsilon'}).$$

Note that for $\epsilon' \in (0, 1]$, $\alpha \epsilon' \triangleq \epsilon \in (0, \alpha]$. This completes the proof. □

Thus, we see that using peaky Gaussian signals, a rate of $C^*(\text{SNR})$ is achievable if the coherence length is greater than or equal to $l^G$.

The required coherence length using peaky Gaussian signaling is strictly greater than the required length (Theorem 1) when using the optimal input distribution. Thus, if $l^*$ is the minimum coherence length needed to have a capacity of $C^*(\text{SNR})$, then

$$l_{\text{min}} < l^* \leq l^G.$$ 

However for $\alpha \in (0, 1]$, as $\epsilon \to 0$, $l^G \to l_{\text{min}}$. Hence, the peaky Gaussian input distribution is near-optimal for the non-coherent MIMO channel. This can be summarized in the following theorem which is a direct consequence of Theorems 1 and 2:

**Theorem 3** Consider a non-coherent Rayleigh block fading channel with average signal to noise ratio $\text{SNR}$. For any $\alpha \in (0, 1]$ and $\epsilon \in (0, \alpha)$, the capacity of the channel is

$$C(\text{SNR}) = r \text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\epsilon})$$

if and only if there exists a $\rho \in (0, \epsilon)$ such that

$$l = \frac{t^2}{(r + t)^2} \text{SNR}^{-2(\alpha+\rho)}.$$ 

This theorem tells us the capacity of a non-coherent MIMO channel in the low SNR regime and shows its dependence on the coherence length of the channel, number of receive and transmit antennas and SNR. Peaky Gaussian signals are near-optimal when communicating over this channel. The theorem leads to the following corollary:

**Corollary 1** Consider a non-coherent Rayleigh block fading channel with average signal to noise ratio $\text{SNR}$. For any $\alpha \in (0, 1]$ and $\epsilon \in (0, \alpha)$, the sublinear capacity term

$$\Delta(l^r)(\text{SNR}) = \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\epsilon})$$

if and only if there exists a $\rho \in (0, \epsilon)$ such that

$$l = \frac{t^2}{(r + t)^2} \text{SNR}^{-2(\alpha+\rho)}.$$ 

21
In Theorem 3 and Corollary 1, \( \alpha \) is used to indicate how close the channel capacity is to the coherent and non-coherent extremes. The coherent channel corresponds to the case when \( \alpha = 1 \) and the i.i.d. non-coherent channel corresponds to the case when \( \alpha \to 0 \). We have also seen that peaky Gaussian signals are optimal for the non-coherent MIMO channel. Thus, with a channel coherence time of \( l \sim \frac{t^2}{(r+t)^2} \text{SNR}^{-2\alpha} \), one should transmit Gaussian signals in \( \delta = \text{SNR}^{1-\alpha} \) fraction of the blocks. At the coherent extreme, \( \delta = \text{SNR}^0 \) and one should transmit in all the blocks in order to achieve capacity. On the other hand, for the i.i.d. Rayleigh fading channel (non-coherent extreme), one should only transmit in \( \delta = \text{SNR}^1 \) fraction of the blocks. We shall study the non-coherent extreme with a finer scaling later on in the paper.

Let us eliminate the parameter \( \alpha \) from Corollary 1. Hence, the sublinear capacity term becomes

\[
\Delta^{(l,r)}(\text{SNR}) = \frac{r}{2\sqrt{l}} \text{SNR} + o\left(\frac{\text{SNR}}{\sqrt{l}}\right).
\]

From (1), we have

\[
\log \left( \frac{E_n}{N_0} \right) \propto \sqrt{\frac{1}{l}}.
\]

Hence, the minimum energy required to transmit and information bit decreases inversely with the square root of the coherence length of the channel. Thus, energy efficiency improves as the coherence length increases. These results apply only for \( \alpha \in (0, 1] \). Increasing the coherence time beyond \( l = \frac{t^2}{(r+t)^2} \text{SNR}^{-2} \) does not decrease the sublinear term significantly. In such cases, the sublinear capacity term remains at the order of \( \text{SNR}^2 \).

We now focus on the coherent and non-coherent extremes.

### 6.3 Coherent Extreme

In this case, \( \alpha = 1 \) and from Theorem 3, we know that for \( \epsilon \in (0, 1) \)

\[
C(\text{SNR}) = r\text{SNR} - \frac{r(r+t)}{2t} \text{SNR}^2 + O(\text{SNR}^{2+\epsilon})
\]

iff there exists a \( \rho \in (0, \epsilon) \) such that

\[
l = \frac{t^2}{(r+t)^2} \text{SNR}^{-2(1+\rho)}. \tag{27}
\]

We see that provided the coherence length is large enough, the non-coherent capacity is the same as the coherent capacity in the low SNR regime. Moreover, the peaky Gaussian signal is now completely continuous. Hence, when \( l \sim \frac{t^2}{(r+t)^2} \text{SNR}^{-2} \), the coherent and non-coherent capacities are the same and continuous Gaussian signals are optimal for both.
6.4 Non-coherent Extreme

From Theorem 3, we see that, as $\alpha \to 0$, $l \to 1$ and we have an i.i.d. Rayleigh fading channel. In order to get the exact value of the sublinear capacity term for this channel, we need to know the precise value of $\alpha$, which is not possible by this asymptotic analysis. We do the precise analysis in Appendix 4 and show that the capacity is

$$C(SNR) = rSNR - \Delta_{i,i.d}^{(t,r)}(SNR).$$

where,

$$\Delta_{i,i.d}^{(t,r)}(SNR) \doteq \frac{rSNR}{\log(SNR^r)}.$$

The notation $\doteq$ denotes an approximation that ignores high order logarithm terms (This approximation is discussed in detail in Appendix 4). The capacity is achieved using on-off input distribution that becomes increasingly “flashy” at low SNR. This is consistent with our asymptotic analysis which showed that only $\delta = SNR^1$ fraction of the blocks should be used for transmission in the non-coherent extreme. Hence, the result shows that besides on-off signaling being optimal for the single input-single output i.i.d. Rayleigh fading channel [11], it is also capacity achieving when multiple antennas are used (with our approximation in mind).

7 Conclusions

In this paper, we consider multiple antenna communication over a wideband, non-coherent Rayleigh block fading channel. With only an average power constraint, we compute the capacity, and consider it’s interaction with the coherence length of the channel, number of transmit and receive antennas and SNR per degree of freedom. This allows us to understand how the energy efficiency of communication and bandwidth penalty are influenced by the number of antennas, coherence length and SNR per degree of freedom. We establish how large the coherence length has to be in order for the non-coherent channel to have a “near coherent” performance in the wideband regime. More specifically, we show that if the coherence length of the channel is above a certain SNR (bandwidth) dependent threshold, the non-coherent and coherent capacities are the same in the wideband regime. We also propose a signaling scheme that in near-optimal.

For the case when the receiver has perfect CSI, we analyze the effect of multiple antennas on the decoding error probability in the infinite bandwidth limit. We show that though multiple transmit antennas do not help from a capacity point of view [12], they do reduce the error probability. A decoding error is caused if either the channel matrix is atypically
ill-conditioned (type 1), which leads to outage, noise is atypically large (type 2), or some codewords are atypically close to each other (type 3). We show that the probability of type 1 errors decreases with increase in the number of transmit and receive antennas, and blocks we code over. The probabilities of type 2 and type 3 errors decreases as we increase the number of receive antennas, coherence length and blocks we code over.
Appendix 1

Proof of Lemma 2:

The conditional probability $p(Y|X,H)$ is given by

$$p(Y|X,H) = \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \exp \left[ - \frac{\text{SNR}}{t} \text{trace}\left\{ (HX - Y)^\dagger(HX - Y) \right\} \right].$$

Define $B$ as

$$B \triangleq \frac{\text{SNR}}{t(1 + \sigma)} H^\dagger H.$$  

$$\Rightarrow B^{-1} = \frac{t(1 + \sigma)}{\text{SNR}} (H^\dagger)^{-1}(H)^{-1}.$$  

We now compute $\int q(X)p(Y|X,H)^{\frac{1}{1+\sigma}} dX$.

$$\int q(X)p(Y|X,H)^{\frac{1}{1+\sigma}} dX$$

$$= \int \frac{1}{\pi^{lt}} \exp \left[ -\text{trace}(X^\dagger X) \right] \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \exp \left[ - \frac{\text{SNR}}{t(1 + \sigma)} \text{trace}\left\{ (HX - Y)^\dagger(HX - Y) \right\} \right] dX$$

$$= \frac{1}{\pi^{lt}} \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \int \exp \left[ -\text{trace}\left\{ X^\dagger (I_t + B) X - \frac{\text{SNR}}{t(1 + \sigma)} (X^\dagger H^\dagger Y + Y^\dagger HX - Y^\dagger Y) \right\} \right] dX$$

$$= \frac{1}{\pi^{lt}} \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \exp \left[ - \frac{\text{SNR}}{t(1 + \sigma)} \text{trace}\left\{ Y^\dagger (I_r - (I_r + B^{-1})^{-1}) Y \right\} \right] \int \exp \left[ - \frac{\text{SNR}}{t(1 + \sigma)} \text{trace}\left\{ X^\dagger H^\dagger (B^{-1} + I_r) H X - X^\dagger H^\dagger Y - Y^\dagger H X + Y^\dagger (I_r + B^{-1})^{-1} Y \right\} \right] dX$$

$$= \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \exp \left[ - \frac{\text{SNR}}{t(1 + \sigma)} \text{trace}\left\{ Y^\dagger (I_r - (I_r + B^{-1})^{-1}) Y \right\} \right] \det(I_t + B)^{-l}$$

Therefore,

$$\int \left[ \int q(X)p(Y|X,H)^{\frac{1}{1+\sigma}} dX \right]^{1+\sigma} dY$$

$$= \left( \frac{\text{SNR}}{\pi t} \right)^{rl} \det(I_t + B)^{-l(1+\sigma)} \int \exp \left[ - \frac{\text{SNR}}{t} \text{trace}\left\{ Y^\dagger (I_r - (I_r + B^{-1})^{-1}) Y \right\} \right] dY$$

$$= \det(I_t + B)^{-l(1+\sigma)} \det \left( I_r - (I_r + B^{-1})^{-1} \right)^{-l}$$

$$= \det(I_t + B)^{-\sigma l}$$

$$= \det \left( I_t + \frac{\text{SNR}}{t(1 + \sigma)} H^\dagger H \right)^{-\sigma l}.$$  

Hence,

$$E_0(\sigma)$$

25
\[- \log E_{H} \left[ \det \left( I_{t} + \frac{\text{SNR}}{t(1 + \sigma)} H^\dagger H \right)^{-\sigma l} \right] \]
\[- \log E_{\Lambda_{1} \ldots \Lambda_{m}} \left[ \prod_{i=1}^{m} \left( 1 + \frac{\text{SNR}}{t(1 + \sigma)} \Lambda_{i} \right)^{-\sigma l} \right] \]
\[\approx - \log E_{\Lambda_{1} \ldots \Lambda_{m}} \left[ \prod_{i=1}^{m} \left( 1 - \frac{\sigma l \text{SNR} \Lambda_{i}}{t(1 + \sigma)} \right) \right] \]
\[= - \log E_{\Lambda_{1} \ldots \Lambda_{m}} \left[ 1 - \frac{\sigma l \text{SNR}}{t(1 + \sigma)} \sum_{i=1}^{m} \Lambda_{i} \right] \]
\[= - \log E_{H} \left[ 1 - \frac{\sigma l \text{SNR}}{t(1 + \sigma)} \text{tr}(H^\dagger H) \right] \]
\[= - \log \left[ 1 - \frac{\sigma l \text{SNR}}{t(1 + \sigma)} \right] \]
\[\approx \frac{\sigma l \text{SNR}}{1 + \sigma}. \tag{32} \]

In (28), \( \Lambda_{i} \) is the \( i \)th unordered positive eigenvalue of the matrix \( H_{k}^\dagger H_{k} \) where \( i \in \{1, \ldots, m\} \) and \( m = \min(t, r) \). Approximations (29), (30) and (32) are valid since we are in the low SNR regime. We obtain (29) since, for small \( x \), \( (1 + x)^{-p} \approx 1 - px \); (30) using the fact that for small \( x_{i}, i \in \{1, \ldots, m\}, \prod_{i=1}^{m} (1 - x_{i}) \approx 1 - \sum_{i=1}^{m} x_{i} \) and, (32) since for small \( x \), \( -\log(1 - x) \approx x \). In (31), we use
\[ E_{H} \left[ \text{tr}(H^\dagger H) \right] = \text{tr}. \]

\[ \square \]

Appendix 2

Proof of Lemma 3:

Proof of (20): For any \( \alpha \in (0, 1] \) and \( \gamma \in (0, \alpha) \), let there exist an input distribution on \( X \) that satisfies the following:
\[ \frac{1}{t} I(X; Y) \geq r \text{SNR} - \frac{r(r + t)}{2t} \text{SNR}^{\dagger+\alpha} + O(\text{SNR}^{\dagger+\alpha+\gamma}). \tag{33} \]

Let \( Y^{G} \) be a matrix, with i.i.d complex Gaussian entries, that satisfies the same power constraint as the received matrix, \( Y \), i.e. \( E[\text{tr}(Y^{G}Y^{G\dagger})] = E[\text{tr}(YY^\dagger)] \). This makes \( h(Y) \leq h(Y^{G}) \) and the entries of \( Y^{G} \) i.i.d \( \mathcal{CN}(0, 1 + \text{SNR}) \). Moreover, conditioned on \( X \), the row vectors of \( Y \) are i.i.d \( \mathcal{CN}(0, XX^\dagger + I_{t}) \). We can thus upper bound the mutual information as
\[ I(X; Y) \]

26
\[ h(Y) - h(Y|X) \leq h(Y^G) - h(Y|X) \]
\[ = rl \log(1 + \text{SNR}) - rtE[\log(1 + \|\tilde{x}_i^T\|^2)] \]
\[ \leq rl\text{SNR} - rtE[\log(1 + \|\tilde{x}_i^T\|^2)] \]  

Combining (33) and (34) and noting that the norms of the input vectors \(\|\tilde{x}_i^T\|\) are identically distributed, we see that if the input distribution satisfies (33), then it necessarily satisfies the first condition (20).

**Proof of (21):** Observing the structure of the optimal input distribution [7] for the noncoherent MIMO channel, we can upper bound the mutual information as

\[ I(X;Y) \leq I(A;Y|\Phi) + I(\Phi;Y|A), \]  

where, \(I(A;Y|\Phi)\) is the information conveyed by the norm of the transmitted signal vectors given that the receiver has side information about their directions, and \(I(\Phi;Y|A)\) is the information conveyed by the direction of these vectors when the receiver has side information about their norm. We establish upper bounds on these two terms.

**Upper bound for \(I(A;Y|\Phi)\):**

When the receiver has side information about \(\Phi\), it can filter out noise orthogonal to the subspace spanned by the row vectors of \(\Phi\) to obtain an equivalent channel

\[ Y\Phi^\dagger =HX\Phi^\dagger + W\Phi^\dagger =HA + W', \]

where \(W'\) has the same distribution as \(W\) and there is no loss in information since \(Y\Phi^\dagger\) is a sufficient statistic for estimating \(X\) from \(Y\). Therefore

\[ I(A;Y|\Phi) = I(A;Y\Phi^\dagger|\Phi) \]
\[ = I(A;HA + W'|\Phi) \]
\[ \leq \sum_{i=1}^{t} I(\|\tilde{x}_i^T\|;h_i\|\tilde{x}_i^T\| + \tilde{w}_i^T|\tilde{\phi}_i^T) \]
\[ \leq \sum_{i=1}^{t} \sum_{j=1}^{r} I(\|\tilde{x}_i^T\|;h_{ij}\|\tilde{x}_i^T\| + w_{ij}'|\tilde{\phi}_i^T), \]  

where the last two inequalities follow from the chain rule of mutual information and the fact that conditioning reduces entropy. In order to get an upper bound on \(I(\|\tilde{x}_i^T\|;h_{ij}\|\tilde{x}_i^T\| + w_{ij}'|\tilde{\phi}_i^T), \)
we need to maximize this mutual information with the average power constraint \( \frac{I}{t} \text{SNR} \) and the constraint specified by (20). If we relax the latter constraint (20), then the mutual information is that of a single-input, single-output i.i.d Rayleigh fading channel with average power constraint \( \frac{I}{t} \text{SNR} \). From [11], we know that this mutual information is maximized by an on-off distribution of the form

\[
||\tilde{x}_i^T||^2 = \begin{cases} \frac{I}{t} \text{SNR} \zeta & \text{w.p. } \zeta \\ 0 & \text{w.p. } 1 - \zeta \end{cases}
\]

for \( \forall i \in \{1, \ldots, t\} \) and some \( \zeta > 0 \). This signalling scheme becomes increasingly “flashy” as the SNR gets low, i.e., \( \zeta \to 0 \) as \( \text{SNR} \to 0 \). Hence, (36) becomes

\[
I(A; Y | \Phi) \\
\leq \sum_{i=1}^{t} \sum_{j=1}^{r} I(||\tilde{x}_i^T||; h_{ij} ||\tilde{x}_i^T|| + w_{ij}' | \phi_i^T) \\
\leq \sum_{i=1}^{t} \sum_{j=1}^{r} H(||\tilde{x}_i^T||) \\
\approx rt \zeta \log(\frac{1}{\zeta}),
\]

where the approximation is valid since we are in the low signal to noise ratio regime and \( \zeta \to 0 \) as \( \text{SNR} \to 0 \). Therefore, we have

\[
\frac{1}{t} I(A; Y | \Phi) \leq \frac{rt \zeta}{t} \log(\frac{1}{\zeta}). \tag{37}
\]

However, this on-off distribution minimizes (20) also and hence the extra constraint does not change the optimal input. Therefore, it suffices to consider on-off signals. Hence, (20) becomes

\[
\frac{(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha + \gamma}) \\
\geq \frac{t \zeta}{t} \log(1 + \frac{t}{\zeta t} \text{SNR}) \\
\geq \frac{t \zeta}{t} \left[ \log(\frac{1}{\zeta}) - \log(\frac{t}{t \text{SNR}}) \right] \\
\approx \frac{t \zeta}{t} \log(\frac{1}{\zeta}), \tag{38}
\]

where the approximation is valid since \( \frac{I}{t} \text{SNR} \gg 1 \) as \( \text{SNR} \to 0 \), i.e., the peak amplitude becomes very large as the signal to noise ratio tends to 0. Combining (37) and (38), we have

\[
\frac{1}{t} I(A; Y | \Phi) \leq \frac{r(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha + \gamma}). \tag{39}
\]
Upper bound for $I(\Phi; Y|A)$:

We can upper bound $I(\Phi; Y|A)$ in terms of the mutual information of a single-input, single-output channel, i.e.

$$I(\Phi; Y|A) \leq \sum_{i=1}^{t} \sum_{j=1}^{r} I(\phi^T_i; \tilde{y}_j^T|A, \phi^T_1, \ldots, \phi^T_{i-1}, \phi^T_{i+1}, \ldots, \phi^T_t, \tilde{y}_1^T, \ldots, \tilde{y}_{j-1}^T, \tilde{y}_{j+1}^T, \ldots, \tilde{y}_r^T).$$

The term inside the double summation represents the mutual information of the channel between the $i^{th}$ transmit antenna and $j^{th}$ receive antenna when no other antenna is present and the norm of $\tilde{x}_i^T$ is known at the receiver. Since the input vectors are identically distributed and the channel matrix has i.i.d. entries, the mutual information between any pair of transmit and receive antennas given that the other antennas are absent will be the same. Hence for all $i \in \{1, \ldots, t\}$ and $j \in \{1, \ldots, r\}$,

$$I(\Phi; Y|A) \leq rt I(\phi^T_i; \tilde{y}_j^T|A, \phi^T_1, \ldots, \phi^T_{i-1}, \phi^T_{i+1}, \ldots, \phi^T_t, \tilde{y}_1^T, \ldots, \tilde{y}_{j-1}^T, \tilde{y}_{j+1}^T, \ldots, \tilde{y}_r^T).$$

We may thus consider the single-input, single-output channel between the $i^{th}$ transmit antenna and $j^{th}$ receive antenna:

$$\tilde{y}_j^T = \mathbf{h}_{ij} ||\tilde{x}_i^T|| + \tilde{w}_j^T.$$

Hence,

$$I(\Phi; Y|A) \leq rt I(\tilde{x}_i^T; \tilde{y}_j^T||\tilde{x}_i^T||) = rt EI(\tilde{x}_i^T; \tilde{y}_j^T||\tilde{x}_i^T||).$$

(40)

Since $I(\tilde{x}_i^T; \tilde{y}_j^T||\tilde{x}_i^T||)$ is the mutual information of a single-input, single-output channel over $l$ channel uses, it has a power constraint of $||\tilde{x}_i^T||^l$. This mutual information can be upper bounded by the capacity of AWGN channel with the same power constraint, i.e.

$$I(\tilde{x}_i^T; \tilde{y}_j^T||\tilde{x}_i^T||) \leq l \log(1 + \frac{||\tilde{x}_i^T||}{l}).$$

(41)

Combining (40) with (41), we obtain

$$I(\Phi; Y|A) \leq rt l \log(1 + \frac{||\tilde{x}_i^T||}{l})$$

(42)

From (47), (39) and (42), we obtain our upper bound to $I(X; Y)$ as

$$\frac{1}{t} I(X; Y) \leq rt l \log(1 + \frac{||\tilde{x}_i^T||}{l}) + \frac{r(r+t)}{2l} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}).$$

(43)

Combining (43) with (33) and noting that all the input vectors have identically distributed norms, we see that the input distribution satisfying (33) satisfies the second constraint (21) also. This completes the proof of the lemma. □

29
Appendix 3

Proof of Theorem 1
For any \(\alpha \in (0, 1]\) and \(\gamma \in (0, \alpha)\), let

\[
C(SNR) \geq C^*(SNR) = rSNR - \frac{r(r + t)}{2t}SNR^{1+\alpha} + O(SNR^{1+\alpha+\gamma}).
\]

This implies that there exists a probability distribution on \(X\) such that

\[
I(X; Y) \geq C^*(SNR).
\]

From Lemma 3, we know that this distribution must satisfy the following constraints for all \(i \in \{1, \ldots, t\}\):

\[
\frac{t}{l} E\left[ \log(1 + \|\mathbf{x}_i^T\|^2) \right] \leq \frac{(r + t)}{2t}SNR^{1+\alpha} + O(SNR^{1+\alpha+\gamma}), \tag{44}
\]

\[
t E\left[ \log \left( 1 + \frac{\|\mathbf{x}_i^T\|^2}{l} \right) \right] \geq SNR - \frac{(r + t)}{l}SNR^{1+\alpha} + O(SNR^{1+\alpha+\gamma}). \tag{45}
\]

Using these constraints, we establish a necessary condition on the coherence length. As the norms of the transmitted signals are identically distributed, it suffices to consider only one of them. Therefore, we will omit the subscript, \(i\), and define random variable \(b\) as

\[
b \triangleq \frac{t\|\mathbf{x}^T\|^2}{lSNR}.
\]

The two constraints become

\[
\frac{t}{l} E\left[ \log \left( 1 + \frac{bSNR}{l} \right) \right] \leq \frac{(r + t)}{2t}SNR^{1+\alpha} + O(SNR^{1+\alpha+\gamma}), \tag{46}
\]

\[
t E\left[ \log \left( 1 + \frac{bSNR}{l} \right) \right] \geq SNR - \frac{(r + t)}{l}SNR^{1+\alpha} + O(SNR^{1+\alpha+\gamma}). \tag{47}
\]

Moreover, as

\[
E[\|\mathbf{x}^T\|^2] = \frac{lSNR}{l}, \tag{48}
\]

\[
\Rightarrow E[b] = 1. \tag{49}
\]

Note that (47,49) do not depend on the coherence length, \(l\), whereas (46) does. Also, the left hand side of (46) is a monotonically decreasing function of \(l\). Thus, to find how large the coherence length necessarily needs to be, we need to find the distribution on \(b\) that minimizes the left hand side of (46) subject to the constraints (47,49). Using this distribution for \(b\), we can obtain the necessary condition on the coherence length from (46).
For any $\beta > 0$, we can express (47) as

$$\frac{\text{SNR}}{t} - \frac{(r + t)}{t^2} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma})$$

$$\leq E\left[\log \left(1 + \frac{\beta \text{SNR}}{t}\right)\right]$$

$$= \Pr(b \geq t\text{SNR}^{-\beta}) E\left[\log \left(1 + \frac{\beta \text{SNR}}{t}\right) \mid b \geq t\text{SNR}^{-\beta}\right]$$

$$+ \Pr(b < t\text{SNR}^{-\beta}) E\left[\log \left(1 + \frac{\beta \text{SNR}}{t}\right) \mid b < t\text{SNR}^{-\beta}\right]$$

$$\leq \Pr(b \geq t\text{SNR}^{-\beta}) E\left[\log \left(1 + \frac{\beta \text{SNR}}{t}\right) \mid b \geq t\text{SNR}^{-\beta}\right] + \Pr(b < t\text{SNR}^{-\beta}) E\left[\frac{\beta \text{SNR}}{t} \mid b < t\text{SNR}^{-\beta}\right]$$

$$= \frac{\text{SNR}}{t} - \Pr(b \geq t\text{SNR}^{-\beta}) E\left[\frac{\beta \text{SNR}}{t} - \log \left(1 + \frac{\beta \text{SNR}}{t}\right) \mid b \geq t\text{SNR}^{-\beta}\right].$$

Therefore,

$$\Pr(b \geq t\text{SNR}^{-\beta}) E\left[\frac{\beta \text{SNR}}{t} - \log \left(1 + \frac{\beta \text{SNR}}{t}\right) \mid b \geq t\text{SNR}^{-\beta}\right] \leq \frac{(r + t)}{t^2} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}).$$

(50)

When $\beta \geq 1$, $b \geq t\text{SNR}^{-\beta}$ implies $\frac{b \text{SNR}}{t} \gg 1$ which makes $\frac{b \text{SNR}}{t} \gg \log(1 + \frac{b \text{SNR}}{t})$. Hence, $\forall \beta \geq 1$

$$\Pr(b \geq t\text{SNR}^{-\beta}) E[b \mid b \geq t\text{SNR}^{-\beta}] \leq \frac{(r + t)}{t^2} \text{SNR}^{\alpha} + O(\text{SNR}^{\alpha+\gamma}) = o(1).$$

(51)

From Markov’s inequality, $\forall \beta \geq 1$

$$\Pr(b \geq t\text{SNR}^{-\beta}) \leq \frac{\text{SNR}^\beta}{t} = o(1).$$

(52)

When $\beta < 1$, $b < t\text{SNR}^{-\beta}$ implies $\frac{b \text{SNR}}{t} \ll 1$. Hence, (50) can be expressed as

$$\frac{(r + t)}{t^2} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma})$$

$$\geq \Pr(b \geq t\text{SNR}^{-\beta}) E\left[\frac{b \text{SNR}}{t} - \log \left(1 + \frac{b \text{SNR}}{t}\right) \mid b \geq t\text{SNR}^{-\beta}\right]$$

$$\geq \Pr \left(t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right) E\left[\frac{b \text{SNR}}{t} - \log \left(1 + \frac{b \text{SNR}}{t}\right) \mid t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right]$$

$$\geq \Pr \left(t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right) E\left[\frac{1}{2} \left(\frac{b \text{SNR}}{t}\right)^2 - \frac{1}{3} \left(\frac{b \text{SNR}}{t}\right)^3 \mid t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right]$$

$$\geq \Pr \left(t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right) \left[\frac{1}{2} \text{SNR}^{2(1-\beta)} - \frac{1}{3} \text{SNR}^{3(1-\beta)}\right].$$

Thus, $\forall \beta \in (0, 1)$

$$\Pr \left(t\text{SNR}^{-1} \geq b \geq t\text{SNR}^{-\beta}\right) \leq \frac{2(r + t)}{t^2} \frac{\text{SNR}^{2\beta-(1-\alpha)}}{1 - \frac{2}{3} \text{SNR}^{1-\beta}} + o\left(\frac{\text{SNR}^{2\beta-(1-\alpha)}}{1 - \frac{2}{3} \text{SNR}^{1-\beta}}\right).$$

(53)
Let us divide the interval \([t \text{SNR}^{-\beta}, t \text{SNR}^{-1}], \beta \in (0, 1),\) into \(K > 1\) finite intervals so that each interval is of length

\[
\nu = \frac{t(\text{SNR}^{-1} - \text{SNR}^{-\beta})}{K}.
\]

Now, for any \(\nu > 0\)

\[
E[b|t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta}] \Pr(t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta})
\]

\[
= \sum_{i=1}^{K} E[b|t \text{SNR}^{\beta+i\nu} \geq b \geq t \text{SNR}^{\beta+(i-1)\nu}] \Pr(t \text{SNR}^{\beta+i\nu} \geq b \geq t \text{SNR}^{\beta+(i-1)\nu})
\]

\[
\leq t \sum_{i=1}^{K} \text{SNR}^{\beta+i\nu} \Pr(t \text{SNR}^{\beta+i\nu} \geq b \geq t \text{SNR}^{\beta+(i-1)\nu})
\]

\[
\leq t \sum_{i=1}^{K} \text{SNR}^{\beta+i\nu} \Pr(t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta})
\]

\[
\leq \frac{2(r+t)}{t} \sum_{i=1}^{K} \text{SNR}^{\beta+i\nu} \left[ \frac{\text{SNR}^{2(\beta+(i+1)\nu) - (1-\alpha)}}{1 - \frac{2}{3} \text{SNR}^{1-(\beta+(i+1)\nu)}} + o\left(\frac{\text{SNR}^{2(\beta+(i+1)\nu) - (1-\alpha)}}{1 - \frac{2}{3} \text{SNR}^{1-(\beta+(i+1)\nu)}}\right)\right]
\]

(54)

Equation (53) is used to obtain (54). Let \(\beta = 1 - \alpha + 2\nu,\) where \(\nu \in (0, \frac{\alpha}{2})\). Thus,

\[
E[b|t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta}] \Pr(t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta})
\]

\[
\leq \frac{2(r+t)}{t} \sum_{i=1}^{K} \left[ \frac{\text{SNR}^{i\nu}}{1 - \frac{2}{3} \text{SNR}^{\alpha-(i+1)\nu}} + o\left(\frac{\text{SNR}^{i\nu}}{1 - \frac{2}{3} \text{SNR}^{\alpha-(i+1)\nu}}\right)\right].
\]

Since,

\[
\frac{\text{SNR}^{i\nu}}{1 - \frac{2}{3} \text{SNR}^{\alpha-(i+1)\nu}} = o(1) \quad \forall i \in \{1, \ldots, K\},
\]

we have

\[
E[b|t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-(1-\alpha+2\nu)}] \Pr(t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-(1-\alpha+2\nu)}) \leq o(1).
\]

(55)

From (53), we have for \(\beta = 1 - \alpha + 2\nu,\) where \(\nu \in (0, \frac{\alpha}{2}),\)

\[
\Pr(t \text{SNR}^{-1} \geq b \geq t \text{SNR}^{-\beta}) \leq \frac{2(r+t)}{t^2} \frac{\text{SNR}^{1-\alpha+4\nu}}{1 - \frac{2}{3} \text{SNR}^{\alpha-2\nu}} + o\left(\frac{\text{SNR}^{1-\alpha+4\nu}}{1 - \frac{2}{3} \text{SNR}^{\alpha-2\nu}}\right) = o(1).
\]

(56)

From (51, 52, 55, 56), we know that for \(0 < \epsilon < \alpha\)

\[
\Pr(b \geq t \text{SNR}^{-(1-\alpha+\epsilon)}) = o(1),
\]

\[
\Pr(b \geq t \text{SNR}^{-(1-\alpha+\epsilon)}) E[b|b \geq t \text{SNR}^{-(1-\alpha+\epsilon)}] = o(1),
\]

(57)
which implies
\[
\Pr(b \leq t \text{SNR}^{-1-\alpha+\epsilon}) = O(1),
\]
\[
\Pr(b \leq t \text{SNR}^{-1-\alpha+\epsilon}) E[b|b \leq t \text{SNR}^{-1-\alpha+\epsilon}] = O(1).
\]

Hence, the distribution on \( b \) that minimizes \( E \left[ \log \left( 1 + \frac{b \text{SNR}}{t} \right) \right] \), subject to the constraints (47,49), is the on-off distribution:
\[
b = \begin{cases} 
t \text{SNR}^{-(1-\alpha+\epsilon)} & \text{w.p. } \eta \\
0 & \text{w.p. } 1 - \eta
\end{cases}
\]
where,
\[
\eta = \frac{\text{SNR}^{1-\alpha+\epsilon}}{t}.
\]

Hence, with this on-off distribution on \( b \), (46) becomes
\[
\frac{(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}) \geq \frac{\text{SNR}^{1-\alpha+\epsilon}}{l} \log \left( 1 + l \text{SNR}^{\alpha-\epsilon} \right).
\]
(57)

Now,
\[
\frac{\text{SNR}^{1-\alpha+\epsilon}}{l} \log \left( 1 + l \text{SNR}^{\alpha-\epsilon} \right) \bigg|_{l = \frac{t^2}{(r+t)^2} \text{SNR}^{-2\alpha+\epsilon}}
\]
\[
= \frac{(r + t)^2}{t^2} \text{SNR}^{1+\alpha} \log \left( 1 + \frac{t^2}{(r+t)^2} \text{SNR}^{-\alpha} \right)
\]
\[
\gg \frac{(r + t)}{2t} \text{SNR}^{1+\alpha} + O(\text{SNR}^{1+\alpha+\gamma}).
\]

Thus, with \( l = \frac{t^2}{(r+t)^2} \text{SNR}^{-2\alpha+\epsilon} \), (57) is not satisfied. However, the right hand side of (57) is a monotonically decreasing function of \( l \). Hence, for the constraint in (57) to be met
\[
l > \frac{t^2}{(r+t)^2} \text{SNR}^{-2\alpha+\epsilon} \quad \forall \epsilon \in (0, \alpha).
\]

Thus, we see that if an input distribution satisfies (44,45,48), then the coherence length must necessarily obey
\[
l > \frac{t^2}{(r+t)^2} \text{SNR}^{-2\alpha} \triangleq l_{\text{min}}.
\]

This completes the proof of the theorem. \( \square \)
Appendix 4

We compute the capacity of an i.i.d Rayleigh fading MIMO channel when CSI is unavailable at both the transmitter as well as the receiver. It is shown in [7] that increasing the number of transmit antennas beyond the coherence length does not increase capacity. Hence, from a capacity point of view, it suffices to have use only one transmit antenna \((t = 1)\). We will therefore consider the capacity of a single-input, multiple-output (SIMO) channel.

Let us pick on-off signaling to communicate over the channel. This signaling scheme is later proved to be optimal for the i.i.d. Rayleigh fading MIMO channel. We specify the signaling as

\[
x = \begin{cases} 
\sqrt{\beta} & \text{w.p. } \lambda \\
0 & \text{w.p. } 1 - \lambda 
\end{cases}
\]

where, \(\beta \in \mathbb{R}^+\) and \(\lambda = \frac{\text{SNR}}{\beta}\). With this signaling, we have the following probability distributions

\[
p_{\tilde{y}|x=0}(\tilde{y}) = \frac{1}{\pi} \exp(-\|\tilde{y}\|^2),
\]

\[
p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y}) = \frac{1}{\pi(1+\beta)} \exp\left(-\frac{\|\tilde{y}\|^2}{1+\beta}\right).
\]

The mutual information \(I(x, \tilde{y})\) can be written as \(I(x, \tilde{y}) = h(\tilde{y}) - h(\tilde{y}|x)\). Now,

\[
h(\tilde{y})
= - \int p_{\tilde{y}}(\tilde{y}) \log(p_{\tilde{y}}(\tilde{y})) d\tilde{y}
= - \int \left[ (1-\lambda) p_{\tilde{y}|x=0}(\tilde{y}) + \lambda p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y}) \right] \log \left[ (1-\lambda) p_{\tilde{y}|x=0}(\tilde{y}) + \lambda p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y}) \right] d\tilde{y}
= -(1-\lambda) \int p_{\tilde{y}|x=0}(\tilde{y}) \log \left[ (1-\lambda) p_{\tilde{y}|x=0}(\tilde{y}) (1 + \frac{\lambda}{1-\lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y})}{p_{\tilde{y}|x=0}(\tilde{y})}) \right] d\tilde{y}
- \lambda \int p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y}) \log \left[ (1-\lambda) p_{\tilde{y}|x=0}(\tilde{y}) (1 + \frac{\lambda}{1-\lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y})}{p_{\tilde{y}|x=0}(\tilde{y})}) \right] d\tilde{y}
= -\log(1-\lambda) - (1-\lambda) \int p_{\tilde{y}|x=0}(\tilde{y}) \log \left[ 1 + \frac{\lambda}{1-\lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y})}{p_{\tilde{y}|x=0}(\tilde{y})} \right] d\tilde{y}
- \lambda \int p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y}) \log \left[ 1 + \frac{\lambda}{1-\lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}}(\tilde{y})}{p_{\tilde{y}|x=0}(\tilde{y})} \right] d\tilde{y} + h(\tilde{y}|x) + \lambda D(p_{\tilde{y}|x=\sqrt{\beta}}||p_{\tilde{y}|x=0})
\]

The divergence \(D(p_{\tilde{y}|x=\sqrt{\beta}}||p_{\tilde{y}|x=0})\) is the divergence between two Gaussian random vectors and is therefore

\[
D(p_{\tilde{y}|x=\sqrt{\beta}}||p_{\tilde{y}|x=0}) = r(\beta - \log(1+\beta))
\]
The expression for the mutual information becomes

$$I(x, \tilde{y}) = r \text{SNR} - r \text{SNR} \frac{\log(1 + \beta)}{\beta} - \log(1 - \frac{\text{SNR}}{\beta}) - I(\text{SNR}, \beta)$$

where,

$$I(\text{SNR}, \beta) = I_1(\text{SNR}, \beta) + I_2(\text{SNR}, \beta), \quad (59)$$

$$I_1(\text{SNR}, \beta) = (1 - \lambda) \int p_{\tilde{y}|x=0}(\tilde{y}) \log \left[ 1 + \frac{\lambda}{1 - \lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}(\tilde{y})}}{p_{\tilde{y}|x=0}(\tilde{y})} \right] d\tilde{y}$$

$$I_2(\text{SNR}, \beta) = \lambda \int p_{\tilde{y}|x=\sqrt{\beta}(\tilde{y})} \log \left[ 1 + \frac{\lambda}{1 - \lambda} \frac{p_{\tilde{y}|x=\sqrt{\beta}(\tilde{y})}}{p_{\tilde{y}|x=0}(\tilde{y})} \right] d\tilde{y}$$

At low SNR, \( \beta \) takes very high values. Therefore, the mutual information can be written as

$$I(x, \tilde{y}) = r \text{SNR} - r \text{SNR} \frac{\log(1 + \beta)}{\beta} - I(\text{SNR}, \beta) \quad (60)$$

We will now compute \( I_1(\text{SNR}, \beta) \) and \( I_2(\text{SNR}, \beta) \). Let us define \( t^* \) to be such that

$$\frac{\text{SNR}}{\beta(1 + \beta)^r} \exp\left(\frac{\beta t^*}{1 + \beta}\right) = 1.$$ 

Note that

$$\frac{t^*}{1 + \beta} = \frac{\log(\beta)}{\beta} + r \frac{\log(1 + \beta)}{\beta} + \frac{\log(\frac{1}{\text{SNR}})}{\beta}.$$ 

Thus,

$$\lim_{\beta \to \infty} \frac{t^*}{1 + \beta} = 0.$$ 

We will use this in future derivations.

### 7.1 Computing \( I_1(\text{SNR}, \beta) \)

$$I_1(\text{SNR}, \beta) = \frac{1}{\pi^r} (1 - \frac{\text{SNR}}{\beta}) \int \exp(-||\tilde{y}||^2) \log \left[ 1 + \frac{\text{SNR}}{(\beta - \text{SNR})(1 + \beta)^r} \exp\left(\frac{\beta}{1 + \beta}||\tilde{y}||^2\right)\right] d\tilde{y}$$

Converting to spherical coordinates in \( 2r \) dimensions, we have for large \( \beta \):

$$I_1(\text{SNR}, \beta)$$

$$= \frac{1}{(r - 1)!} \int_0^\infty t^{r-1} \exp(-t) \log \left[ 1 + \frac{\text{SNR}}{\beta(1 + \beta)^r} \exp\left(\frac{\beta t}{1 + \beta}\right)\right] dt + o(\text{SNR}^2)$$

$$= \frac{\exp(-t^*)}{(r - 1)!} \int_0^\infty t^{r-1} \exp(-t - t^*) \log \left[ 1 + \exp\left(\frac{\beta(t - t^*)}{1 + \beta}\right)\right] dt + o(\text{SNR}^2)$$

$$= \frac{\exp(-t^*)}{(r - 1)!} \int_{-t^*}^\infty (t + t^*)^{r-1} \exp(-t) \log \left[ 1 + \exp\left(\frac{\beta t}{1 + \beta}\right)\right] dt + o(\text{SNR}^2)$$

$$= I_{11}(\text{SNR}, \beta) + I_{12}(\text{SNR}, \beta) + o(\text{SNR}^2), \quad (61)$$
where,

\[
I_{11}(\text{SNR}, \beta) = \frac{\exp(-t^\ast)}{(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( \frac{\beta t}{1+\beta} \right) \right] dt
\]

\[
I_{12}(\text{SNR}, \beta) = \frac{\exp(-t^\ast)}{(r-1)!} \int_{0}^{\infty} (t + t^\ast)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( \frac{\beta t}{1+\beta} \right) \right] dt
\]

7.1.1 Computing \( I_{11}(\text{SNR}, \beta) \)

\[
I_{11}(\text{SNR}, \beta)
= \frac{\exp(-t^\ast)}{(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( \frac{\beta t}{1+\beta} \right) \right] dt
\]

\[
\leq \frac{\exp(-t^\ast)}{(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) \exp \left( \frac{\beta t}{1+\beta} \right) dt
\]

\[
\leq \frac{\exp(-t^\ast)}{(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) dt
\]

\[
= \frac{1 + \beta^{r}}{(r-1)!} \exp(-\frac{\beta t^\ast}{1+\beta}) \int_{0}^{t^\ast} t^{r-1} \exp(-t) dt
\]

\[
= \frac{1 + \beta^{r}}{(r-1)!} \exp(-\frac{\beta t^\ast}{1+\beta}) \Gamma(r) - \Gamma(r, \frac{t^\ast}{1+\beta})
\]

\[
= I_{11}^{U}(\text{SNR}, \beta),
\]

where,

\[
I_{11}^{U}(\text{SNR}, \beta) = \frac{\text{SNR}}{\beta} - (1 + \beta)^r \left[ \frac{\text{SNR}}{\beta(1+\beta)^r} \right]^{1+\frac{r}{2}} \left[ \sum_{j=0}^{r-1} \frac{\Gamma(r-1)}{j!} \right]
\]

Moreover,

\[
I_{11}(\text{SNR}, \beta)
= \frac{\exp(-t^\ast)}{(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( \frac{\beta t}{1+\beta} \right) \right] dt
\]

\[
\geq I_{11}^{U}(\text{SNR}, \beta) - \frac{\exp(-t^\ast)}{2(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp(-t) \exp \left( \frac{2\beta t}{1+\beta} \right) dt
\]

\[
= I_{11}^{U}(\text{SNR}, \beta) - \frac{\exp(-t^\ast)}{2(r-1)!} \int_{-t^\ast}^{0} (t + t^\ast)^{r-1} \exp \left( \frac{\beta - 1}{2\beta + 1} t \right) dt
\]

\[
= I_{11}^{U}(\text{SNR}, \beta) - \frac{\exp(-t^\ast)}{2(r-1)!} \int_{0}^{t^\ast} t^{r-1} \exp \left( \frac{\beta - 1}{2\beta + 1} (t - t^\ast) \right) dt
\]

\[
= I_{11}^{U}(\text{SNR}, \beta) - \frac{1}{2(r-1)!} \exp \left( - \frac{2\beta t^\ast}{\beta + 1} \right) \left[ \frac{\beta + 1}{\beta - 1} \right]^{r} \int_{0}^{(\frac{\beta + 1}{\beta - 1})t^\ast} t^{r-1} \exp(t) dt
\]

36
\[
I_{11}(\text{SNR}, \beta) - \frac{(-1)^{r-1}}{2(r-1)!} \exp \left( -\frac{2\beta t^*}{\beta + 1} \right) \left[ \frac{\beta + 1}{\beta - 1} \right]^r \left[ \Gamma\left(r, -\frac{\beta - 1}{\beta + 1} t^*\right) - \Gamma(r) \right]
\]

\[
= I_{11}(\text{SNR}, \beta) - \frac{(-1)^{r-1}}{2} \left[ \frac{\beta + 1}{\beta - 1} \right]^r \left[ \exp(-t^*) \sum_{j=0}^{r-1} \frac{[-\frac{\beta - 1}{\beta + 1} t^*]^j}{j!} - \exp \left( -\frac{2\beta t^*}{\beta + 1} \right) \right]
\]

\[
= I_{11}(\text{SNR}, \beta) - \frac{(-1)^{r-1}}{2} \left[ \frac{\beta + 1}{\beta - 1} \right]^r \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1+\frac{1}{\beta}} \left[ \sum_{j=0}^{r-1} \frac{[-\frac{\beta - 1}{\beta + 1} t^*]^j}{j!} \right] + o(\text{SNR}^2)
\]

Let

\[
I_1^L(\text{SNR}, \beta) = \frac{(-1)^{r-1}}{2} \left[ \frac{\beta + 1}{\beta - 1} \right]^r \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1+\frac{1}{\beta}} \left[ \sum_{j=0}^{r-1} \frac{[-\frac{\beta - 1}{\beta + 1} t^*]^j}{j!} \right]
\]

Thus, we have

\[
I^U_{11}(\text{SNR}, \beta) - I^L_{11}(\text{SNR}, \beta) + o(\text{SNR}^2) \leq I_{11}(\text{SNR}, \beta) \leq I^U_{11}(\text{SNR}, \beta)
\]

Since \( \frac{\beta}{1+\beta} \to 0 \) as \( \beta \to \infty \), we have

\[
\lim_{\beta \to \infty} I^U_{11}(\text{SNR}, \beta) = 0,
\]

\[
\lim_{\beta \to \infty} I^L_{11}(\text{SNR}, \beta) = 0,
\]

\[
\Rightarrow \lim_{\beta \to \infty} I_{11}(\text{SNR}, \beta) = 0.
\]

(62)

7.1.2 Computing \( I_{12}(\text{SNR}, \beta) \)

\[
I_{12}(\text{SNR}, \beta) = \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t + t^*)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( \frac{\beta t}{1 + \beta} \right) \right] dt
\]

\[
= \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t + t^*)^{r-1} \exp(-t) \left[ \frac{\beta t}{1 + \beta} + \log \left[ 1 + \exp \left( -\frac{\beta t}{1 + \beta} \right) \right] \right] dt
\]

\[
= I^1_{12}(\text{SNR}, \beta) + I^2_{12}(\text{SNR}, \beta)
\]

(63)

where,

\[
I^1_{12}(\text{SNR}, \beta) = \frac{\exp(-t^*)}{(r-1)!} \left[ \frac{\beta}{1 + \beta} \right] \int_0^\infty t(t + t^*)^{r-1} \exp(-t) dt,
\]

\[
I^2_{12}(\text{SNR}, \beta) = \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t + t^*)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( -\frac{\beta t}{1 + \beta} \right) \right] dt.
\]

37
Now,

\[ I_{12}^1(\text{SNR}, \beta) \]

\[
= \frac{\exp(-t^*)}{(r-1)!} \left[ \frac{\beta}{1+\beta} \right] \int_0^\infty t(t+t^*)^{r-1} \exp(-t) dt,
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{\beta}{1+\beta} \right] \int_{t^*}^\infty (t-t^*)^{r-1} \exp(-t) dt,
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{\beta}{1+\beta} \right] \Gamma(r+1, t^*) - t^* \Gamma(r, t^*)
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{\beta}{1+\beta} \right] \Gamma(r, t^*)(r-t^*) + (t^*)^r \exp(-t^*)
\]

\[
= \exp(-t^*) \left[ \frac{\beta}{1+\beta} \right] \left[ \frac{r-t^*}{(r-1)!} \left[ \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] + \frac{(t^*)^r}{(r-1)!} \right]
\]

\[
= \left[ \frac{\text{SNR}}{(\beta(1+\beta)^r)} \right]^{1+\frac{1}{r}} \left[ \frac{\beta}{1+\beta} \right] \left[ \frac{r-t^*}{(r-1)!} \left[ \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] + \frac{(t^*)^r}{(r-1)!} \right]
\]

\[
= I_{12}^{1A}(\text{SNR}, \beta) \text{SNR}^{1+\frac{1}{r}}
\]

where,

\[
I_{12}^{1A}(\text{SNR}, \beta) = \left[ \frac{1}{\beta(1+\beta)^r} \right]^{1+\frac{1}{r}} \left[ \frac{\beta}{1+\beta} \right] \left[ \frac{r-t^*}{(r-1)!} \left[ \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] + \frac{(t^*)^r}{(r-1)!} \right]
\]

Since \( \frac{t^*}{\beta+1} \to 0 \) as \( \beta \to \infty \), we have

\[
\lim_{\beta \to \infty} I_{12}^{1A}(\text{SNR}, \beta) = 0.
\]

Thus,

\[
\lim_{\beta \to \infty} I_{12}^1(\text{SNR}, \beta) = 0.
\] (64)

We will now compute \( I_{12}^2(\text{SNR}, \beta) \).

\[
I_{12}^2(\text{SNR}, \beta)
\]

\[
= \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t+t^*)^{r-1} \exp(-t) \log \left[ 1 + \exp \left( -\frac{\beta t}{1+\beta} \right) \right] dt
\]

\[
\leq \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t+t^*)^{r-1} \exp(-t) \exp \left( -\frac{\beta t}{1+\beta} \right) dt
\]

\[
= \frac{\exp(-t^*)}{(r-1)!} \int_0^\infty (t+t^*)^{r-1} \exp\left( -\frac{1+2\beta}{1+\beta} t \right) dt
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{1+\beta}{1+2\beta} \right] \exp \left( \frac{\beta t}{1+\beta} \right) \int_{(\frac{1+2\beta}{1+\beta})t^*}^\infty t^{r-1} \exp(-t) dt
\]

38
\[
\frac{1}{(r-1)!} \left[ \frac{1 + \beta}{1 + 2\beta} \right]^r \exp \left( \frac{\beta t}{1 + \beta} \right) \Gamma(r, \left( \frac{1 + 2\beta}{1 + \beta} \right)t^*) \\
= \left[ \frac{1 + \beta}{1 + 2\beta} \right]^r \exp(-t^*) \sum_{j=0}^{r-1} \frac{\left( \frac{1 + 2\beta}{1 + \beta} \right)^j t^*}{j!} \\
= I_{12}^{2U}(\text{SNR}, \beta),
\]

where,
\[
I_{12}^{2U}(\text{SNR}, \beta) = \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1 + \frac{1}{r}} \left[ \frac{1 + \beta}{1 + 2\beta} \right]^r \left[ \sum_{j=0}^{r-1} \frac{\left( \frac{1 + 2\beta}{1 + \beta} \right)^j t^*}{j!} \right].
\]

Moreover,
\[
I_{12}^{2L}(\text{SNR}, \beta)
\]
\[
\geq I_{12}^{2U}(\text{SNR}, \beta) - \frac{\exp(-t^*)}{2(r-1)!} \int_0^\infty (t + t^*)^{r-1} \exp(-t) \exp \left( - \frac{2\beta t}{1 + \beta} \right) dt \\
= I_{12}^{2U}(\text{SNR}, \beta) - \frac{1}{2(r-1)!} \exp \left( \frac{2\beta t}{1 + \beta} \right) \int_{t^*}^\infty t^{r-1} \exp(-\left( \frac{1 + 3\beta}{1 + \beta} \right)t) dt \\
= I_{12}^{2U}(\text{SNR}, \beta) - \frac{1}{2(r-1)!} \left[ \frac{1 + \beta}{1 + 3\beta} \right]^r \exp \left( \frac{2\beta t}{1 + \beta} \right) \Gamma(r, \left( \frac{1 + 3\beta}{1 + \beta} \right)t^*) \\
= I_{12}^{2U}(\text{SNR}, \beta) - \frac{\exp(-t^*)}{2} \left[ \frac{1 + \beta}{1 + 3\beta} \right]^r \left[ \sum_{j=0}^{r-1} \frac{\left( \frac{1 + 3\beta}{1 + \beta} \right)^j t^*}{j!} \right] \\
= I_{12}^{2U}(\text{SNR}, \beta) - \frac{1}{2} I_{12}^{2L}(\text{SNR}, \beta).
\]

where,
\[
I_{12}^{2L}(\text{SNR}, \beta) = \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1 + \frac{1}{r}} \left[ \frac{1 + \beta}{1 + 2\beta} \right]^r \left[ \sum_{j=0}^{r-1} \frac{\left( \frac{1 + 3\beta}{1 + \beta} \right)^j t^*}{j!} \right].
\]

Since \( \frac{t^*}{1 + \beta} \to 0 \) as \( \beta \to \infty \),
\[
\lim_{\beta \to \infty} I_{12}^{2L}(\text{SNR}, \beta) = 0, \\
\lim_{\beta \to \infty} I_{12}^{2U}(\text{SNR}, \beta) = 0, \\
\Rightarrow \lim_{\beta \to \infty} I_{12}^{2U}(\text{SNR}, \beta) = 0.
\]

Substituting (64) and (65) in (63), we have
\[
\lim_{\beta \to \infty} I_{12}(\text{SNR}, \beta) = 0.
\]  

(66)

Substituting (62) and (66) in (61), we obtain
\[
I_1(\text{SNR}, \beta) = o(\text{SNR}^2).
\]

(67)
7.2 Computing $I_2(SNR, \beta)$

$$I_2(SNR, \beta) = \frac{SNR}{\pi^{\frac{r}{2}}(1 + \beta)^{r}} \int \exp(- \frac{||\tilde{y}||^2}{2(1 + \beta)}) \log[1 + \frac{SNR}{(\beta - SNR)(1 + \beta)^{r}} \exp(\frac{\beta t}{1 + \beta})]d\tilde{y}$$

Converting to spherical coordinates in $2r$ dimensions, we have for large $\beta$:

$$I_2(SNR, \beta) = \frac{SNR}{\beta(\beta + 1)^r(r - 1)!} \int_0^\infty t^{r-1} \exp(- t \frac{1}{1 + \beta}) \log[1 + \frac{SNR}{\beta(1 + \beta)^r} \exp(\frac{\beta t}{1 + \beta})]dt$$

$$= I_{21}(SNR, \beta) + I_{22}(SNR, \beta) + o(SNR^2), \quad (68)$$

where

$$I_{21}(SNR, \beta) = \frac{SNR}{\beta(\beta + 1)^r(r - 1)!} \int_0^t t^{r-1} \exp(- t \frac{1}{1 + \beta}) \log[1 + \frac{SNR}{\beta(1 + \beta)^r} \exp(\frac{\beta t}{1 + \beta})]dt$$

$$I_{22}(SNR, \beta) = \frac{SNR}{\beta(\beta + 1)^r(r - 1)!} \int_t^\infty t^{r-1} \exp(- t \frac{1}{1 + \beta}) \log[1 + \frac{SNR}{\beta(1 + \beta)^r} \exp(\frac{\beta t}{1 + \beta})]dt. \quad (68)$$

7.2.1 Computing $I_{21}(SNR, \beta)$

$$I_{21}(SNR, \beta) = \frac{SNR}{\beta(\beta + 1)^r(r - 1)!} \int_0^t t^{r-1} \exp(- t \frac{1}{1 + \beta}) \log \left[1 + \frac{SNR}{\beta(1 + \beta)^r} \exp \left(\frac{\beta t}{1 + \beta}\right)\right]dt$$

$$\leq \frac{SNR}{\beta(\beta + 1)^r(r - 1)!} \int_0^t t^{r-1} \exp \left(- t \frac{1}{1 + \beta}\right) \frac{SNR}{\beta(1 + \beta)^r} \exp \left(\frac{\beta t}{1 + \beta}\right)dt$$

$$= \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^2 \frac{1}{(r - 1)!} \int_0^t t^{r-1} \exp \left(\frac{\beta - 1}{\beta + 1} t\right)dt$$

$$= \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^2 \left[\frac{\beta + 1}{\beta - 1}\right]^r \frac{1}{(r - 1)!} \int_0^t t^{r-1} \exp(t)dt$$

$$= \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^2 \left[\frac{\beta + 1}{\beta - 1}\right]^r \frac{(-1)^{r-1}}{(r - 1)!} \left[\Gamma(r, -\left(\frac{\beta - 1}{\beta + 1}\right) t) - \Gamma(r)\right]$$

$$= \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^2 \left[\frac{\beta + 1}{\beta - 1}\right]^r \frac{(-1)^{r-1}}{(r - 1)!} \left[(r - 1)! \exp \left(\frac{\beta - 1}{\beta + 1}\right) t^* \right] \sum_{j=0}^{r-1} \frac{(-1)^{j+1} t^*}{j!} - (r - 1)!$$

$$= \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^2 \left[\frac{\beta + 1}{\beta - 1}\right]^r \frac{(-1)^{r-1}}{(r - 1)!} \left[(r - 1)! \left[\frac{SNR}{\beta(1 + \beta)^r}\right]^\frac{1}{2} \sum_{j=0}^{r-1} \frac{(-1)^{j+1} t^*}{j!} - (r - 1)!\right]$$

$$= [I_{21}^U(SNR, \beta)]^2 SNR^{1+\frac{1}{2}} + o(SNR^2)$$

40
where,

\[
I_{21}^U(SNR, \beta) = \frac{(-1)^r}{[\beta(1 + \beta)^r]^{1 + \frac{1}{\beta}} [2\beta - 1]} \sum_{j=0}^{r-1} \frac{[-(\frac{2\beta-1}{\beta+1})t^*]^j}{j!}
\]

Now,

\[
I_{21}(SNR, \beta) = \frac{SNR}{\beta(1 + \beta)^r} \int_0^t \frac{t^{r-1}}{(r-1)!} \exp \left( -\frac{t}{1 + \beta} \right) \log[1 + \frac{SNR}{\beta(1 + \beta)^r} \exp \left( \frac{\beta t}{1 + \beta} \right)] dt
\]

\[
\geq \frac{SNR}{2\beta(1 + \beta)^r} \int_0^t \frac{t^{r-1}}{(r-1)!} \exp \left( -\frac{t}{1 + \beta} \right) \left[ \frac{SNR}{\beta(1 + \beta)^r} \exp \left( \frac{\beta t}{1 + \beta} \right) \right]^2 dt
\]

\[
= \left[ I_{21}^U(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

\[
- \frac{1}{2} \left[ \frac{SNR}{\beta(1 + \beta)^r} \right]^3 \frac{1}{(r-1)!} \int_0^t \frac{t^{r-1}}{(r-1)!} \exp \left( \frac{2\beta-1}{1 + \beta} t \right) dt
\]

\[
= \left[ I_{21}^U(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

\[
- \frac{1}{2} \left[ \frac{SNR}{\beta(1 + \beta)^r} \right]^3 \frac{(-1)^r}{(r-1)!} \left[ \frac{\beta + 1}{2\beta - 1} \right] \left[ \Gamma(r, -(\frac{2\beta-1}{1 + \beta})t^*) - \Gamma(r) \right]
\]

\[
= \left[ I_{21}^U(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

\[
- \frac{1}{2} \left[ \frac{SNR}{\beta(1 + \beta)^r} \right]^3 \frac{(-1)^r}{(r-1)!} \left[ \frac{\beta + 1}{2\beta - 1} \right] \left[ (r-1)! \left\{ \frac{\beta(1 + \beta)^r}{\beta(1 + \beta)^r} \right\} \right] \sum_{j=0}^{\frac{r-1}{2}} \frac{[-(\frac{2\beta-1}{\beta+1})t^*]^j}{j!} - (r-1)!
\]

\[
= \left[ I_{21}^U(SNR, \beta) - \frac{1}{2} I_{21}^L(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

where,

\[
I_{21}^L(SNR, \beta) = \frac{(-1)^r}{[\beta(1 + \beta)^r]^{1 + \frac{1}{\beta}} [2\beta - 1]} \sum_{j=0}^{r-1} \frac{[-(\frac{2\beta-1}{\beta+1})t^*]^j}{j!}
\]

Combining the lower and upper bounds, we have

\[
\left[ I_{21}^U(SNR, \beta) - \frac{1}{2} I_{21}^L(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

\[
\leq I_{21}(SNR, \beta) \leq \left[ I_{21}^U(SNR, \beta) \right] SNR^{1 + \frac{1}{\beta}} + o(SNR^2)
\]

At low SNR, \( \frac{t^*}{1 + \beta} \to 0 \) as \( \beta \to \infty \). Thus, we have

\[
\lim_{\beta \to \infty} I_{21}^U(SNR, \beta) = 0
\]
\[
\lim_{\beta \to \infty} I_{21}^L(\text{SNR}, \beta) = 0
\]

Therefore,
\[
I_{21}(\text{SNR}, \beta) = o(\text{SNR}^2). \tag{69}
\]

### 7.2.2 Computing \(I_{22}(\text{SNR}, \beta)\)

\[
I_{22}(\text{SNR}, \beta)
= \frac{\text{SNR}}{\beta(\beta + 1)^r} \int_{t^*}^{\infty} t^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) \log \left[ 1 + \frac{\text{SNR}}{\beta(1 + \beta)^r} \exp \left( \frac{\beta t}{1 + \beta} \right) \right] dt
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1 + \frac{1}{r}} \int_{t^*}^{\infty} t^{r-1} \exp \left( - \frac{(t-t^*)}{1 + \beta} \right) \log \left[ 1 + \exp \left( \frac{\beta(t-t^*)}{1 + \beta} \right) \right] dt
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{\text{SNR}}{\beta(1 + \beta)^r} \right]^{1 + \frac{1}{r}} \int_{t^*}^{\infty} (t+t^*)^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) \log \left[ 1 + \exp \left( \frac{\beta t}{1 + \beta} \right) \right] dt
\]

\[
= \frac{1}{(r-1)!} \left[ \frac{1}{\beta(1 + \beta)^r} \right]^{1 + \frac{1}{r}} \int_{0}^{t^*} (t+t^*)^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) \log \left[ 1 + \exp \left( \frac{\beta t}{1 + \beta} \right) \right] dt\tag{70}
\]

where,
\[
I_{22}^1(\text{SNR}, \beta) = \frac{1}{(1 + \beta)^{r+1}} \int_{0}^{\infty} t(t+t^*)^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) dt,
\]

\[
I_{22}^2(\text{SNR}, \beta) = \frac{1}{\beta(1 + \beta)^r} \int_{0}^{\infty} (t+t^*)^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) \log \left[ 1 + \exp \left( - \frac{\beta t}{1 + \beta} \right) \right] dt.
\]

Now,
\[
I_{22}^1(\text{SNR}, \beta)
= \frac{1}{(1 + \beta)^{r+1}} \int_{0}^{\infty} t(t+t^*)^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) dt
\]

\[
= \frac{1}{(1 + \beta)^{r+1}} \exp \left( \frac{t^*}{1 + \beta} \right) \left[ \int_{t^*}^{\infty} t^{r} \exp \left( - \frac{t}{1 + \beta} \right) dt - t^* \int_{t^*}^{\infty} t^{r-1} \exp \left( - \frac{t}{1 + \beta} \right) dt \right]
\]

\[
= \frac{1}{(1 + \beta)^{r+1}} \exp \left( \frac{t^*}{1 + \beta} \right) \left[ (1 + \beta)^{r+1} \Gamma(r+1, \frac{t^*}{1 + \beta}) - t^*(1 + \beta)^r \Gamma(r, \frac{t^*}{1 + \beta}) \right]
\]

\[
= \frac{1}{(1 + \beta)^{r+1}} \exp \left( \frac{t^*}{1 + \beta} \right) \left[ (r+1 + \beta)^{r+1} \Gamma(r, \frac{t^*}{1 + \beta}) + (1 + \beta)^{r+1} \Gamma(r, \frac{t^*}{1 + \beta}) - t^* \Gamma(r, \frac{t^*}{1 + \beta}) \right]
\]

\[
= \frac{\beta}{1 + \beta} I_{22}^{1A}(r, \text{SNR}, \beta),
\]
where,

\[ I_{22}^{1A}(r, \text{SNR}, \beta) = r! \left[ \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] \left[ 1 - \frac{t^*}{r(1+\beta)} \right] + \left[ \frac{t^*}{1+\beta} \right]^r. \]

Since \( \frac{t^*}{1+\beta} \to 0 \) as \( \beta \to \infty \),

\[ \lim_{\beta \to \infty} I_{22}^{1A}(r, \text{SNR}, \beta) = r! \] \hspace{1cm} (71)

Thus,

\[ I_{22}^{1}(r, \text{SNR}, \beta) = \frac{\beta r!}{1 + \beta}. \] \hspace{1cm} (72)

We now compute \( I_{22}^{2}(\text{SNR}, \beta) \).

\[
\begin{align*}
I_{22}^{2}(\text{SNR}, \beta) &= \frac{1}{\beta(1+\beta)^r} \int_0^\infty (t + t^*)^{r-1} \exp \left( - \frac{t}{1+\beta} \right) \log \left[ 1 + \exp \left( - \frac{\beta t}{1+\beta} \right) \right] dt \\
&\leq \frac{1}{\beta(1+\beta)^r} \int_0^\infty (t + t^*)^{r-1} \exp \left( - \frac{t}{1+\beta} \right) \exp \left( - \frac{\beta t}{1+\beta} \right) dt \\
&= \frac{1}{\beta(1+\beta)^r} \exp(t^*) \int_{t^*}^\infty t^{r-1} \exp(-t) dt \\
&= \frac{1}{\beta(1+\beta)^r} \exp(t^*) \Gamma(r, t^*) \\
&= \frac{1}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{1}{j!} (t^*)^j \\
&= \frac{1}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \\
&= \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!}.
\end{align*}
\]

Moreover,

\[
\begin{align*}
I_{22}^{2}(\text{SNR}, \beta) \\
&\geq \frac{1}{\beta(1+\beta)^r} \int_0^\infty (t + t^*)^{r-1} \exp \left( - \frac{t}{1+\beta} \right) \exp \left( - \frac{\beta t}{1+\beta} \right) dt \\
&\geq \frac{1}{2\beta(1+\beta)^r} \int_0^\infty (t + t^*)^{r-1} \exp \left( - \frac{t}{1+\beta} \right) \exp \left( - \frac{2\beta t}{1+\beta} \right) dt \\
&= \left[ \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] - \frac{1}{2\beta(1+\beta)^r} \int_0^\infty (t + t^*)^{r-1} \exp \left( - \frac{(1+2\beta)t}{1+\beta} \right) dt \\
&= \left[ \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] - \frac{1}{2\beta(1+\beta)^r} \exp \left( \frac{(1+2\beta)t^*}{1+\beta} \right) \left[ \frac{1+\beta}{1+2\beta} \right]^r \Gamma(r, \frac{(1+2\beta)t^*}{1+\beta}) \\
&= \left[ \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] - \frac{(r-1)!}{2\beta(1+\beta)^r} \left[ \frac{1+\beta}{1+2\beta} \right]^r \sum_{j=0}^{r-1} \frac{(1+2\beta)j^*}{j!}
\end{align*}
\]
Now as $\beta \to \infty$, $\frac{\nu}{1+\beta} \to 0$. Therefore,

$$\lim_{\beta \to \infty} \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} = 0,$$

$$\lim_{\beta \to \infty} \left[ \frac{(r-1)!}{\beta(1+\beta)^r} \sum_{j=0}^{r-1} \frac{(t^*)^j}{j!} \right] - \frac{(r-1)!}{2\beta(1+\beta)^r} \left[ \frac{1 + \beta}{1 + 2\beta} \right] \sum_{j=0}^{r-1} \frac{[(1+\frac{2\beta}{1+\beta})t^*]^j}{j!} = 0.$$

As both the upper and lower bounds go to 0, we have

$$\lim_{\beta \to \infty} I_{22}(\text{SNR}, \beta) = 0.$$ (73)

Substituting (72) and (73) in (70), we have

$$I_{22}(\text{SNR}, \beta) = \frac{1}{(r-1)!} \left[ \frac{1}{\beta(1+\beta)^r} \right]^\frac{1}{2} \text{SNR}^{1+\frac{1}{2}} \left[ \frac{\beta r!}{1+\beta} \right] + o(\text{SNR}^2),$$

$$= r \beta^{-\frac{r+1}{\beta}} \text{SNR}^{1+\frac{1}{2}} + o(\text{SNR}^2).$$

Therefore,

$$I_2(\text{SNR}, \beta) = r \beta^{-\frac{r+1}{\beta}} \text{SNR}^{1+\frac{1}{2}} + o(\text{SNR}^2).$$ (74)

Substituting (67) and (74) in (60) and (59), we obtain

$$I(x, \bar{y}) = r \text{SNR} - r \text{SNR} \frac{\log(1+\beta)}{\beta} - r \beta^{-\frac{r+1}{\beta}} \text{SNR}^{1+\frac{1}{2}} + o(\text{SNR}^2).$$

Let the capacity of the channel be $C(\text{SNR})$. Since, on-off signaling may not be optimal for the channel, we will denote the highest achievable rate using on-off signaling as $C_{\text{on-off}}(\text{SNR})$. $C_{\text{on-off}}(\text{SNR})$, is given by

$$C_{\text{on-off}}(\text{SNR}) = \max_{\beta} I(x, \bar{y})$$

$$= r \text{SNR}[1 - M^*(\text{SNR})] + o(\text{SNR}^2),$$ (75)

where,

$$M^*(\text{SNR}) = \min_{\beta} \left[ \frac{\log(1+\beta)}{\beta} + \beta^{-\frac{r+1}{\beta}} \text{SNR}^{\frac{1}{2}} \right]$$

$$= \min_{\beta} \left[ \frac{-1}{\beta} + \beta^{-\frac{r+1}{\beta}} \text{SNR}^{\frac{1}{2}} \right].$$

44
The last equality holds since $\beta$ is large. Let us denote

$$M(\beta, \text{SNR}) = \min_{\beta} \left[ \frac{\log(\beta)}{\beta} + \beta^{-\frac{r+1}{\text{SNR}}} \right].$$

We will prove the following theorem to get a lower bound on $M^*(\text{SNR})$.

$$M^*(\text{SNR}) \geq M_L(\text{SNR}) \triangleq \frac{\log \log(\frac{r}{\text{SNR}})}{\log(\frac{r}{\text{SNR}})} \tag{76}$$

We will prove this by contradiction. Let there be an $\beta_1$ such that the theorem does not hold. Since $\frac{\log(\beta_1)}{\beta_1} \geq 0$ and $\beta_1^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \geq 0$, we have,

$$\frac{\log(\beta_1)}{\beta_1} \leq M_L(\text{SNR}), \quad (77)$$

$$\beta_1^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \leq M_L(\text{SNR}). \quad (78)$$

If (77) holds, we have

$$\beta_1 \geq \log\left(\frac{r}{\text{SNR}}\right).$$

Moreover,

$$\beta_1^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \geq \beta_1^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \left[ \frac{\text{SNR}}{r} \right]^{\frac{1}{2}}
= \exp\left(-\frac{(r+1) \log(\beta_1)}{\beta_1}\right) \left[ \frac{\text{SNR}}{r} \right]^{-\frac{1}{2} \log(\frac{r}{\text{SNR}})}
\geq \exp[-(r+1)M_L(\text{SNR})]e^{-1}$$

As $\text{SNR} \to 0$, we have

$$\exp[-(r+1)M_L(\text{SNR})]e^{-1} \gg M_L(\text{SNR}),$$

$$\Rightarrow \beta_1^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \gg M_L(\text{SNR}).$$

This contradicts (78), which completes the proof. To get an upper bound for $M^*(\text{SNR})$, we pick a value of $\beta$. Let

$$\beta_2 = \frac{\log(\frac{r}{\text{SNR}})}{\log \log(\frac{r}{\text{SNR}})}.$$

Now,

$$M^*(\text{SNR}) \leq \frac{\log(\beta_2)}{\beta_2} + \beta_2^{-\frac{r+1}{\text{SNR}^{\frac{1}{2}}}} \tag{79}.$$
We have
\[
\begin{align*}
\frac{\log(\beta_2)}{\beta_2} & = \frac{[\log \log(\frac{r}{\text{SNR}}) - \log \log(\frac{r}{\text{SNR}})] \log(\frac{r}{\text{SNR}})}{\log(\frac{r}{\text{SNR}})} \\
& \leq \frac{[\log(\frac{r}{\text{SNR}})]^2}{\log(\frac{r}{\text{SNR}})}, \quad (80)
\end{align*}
\]
and,
\[
\begin{align*}
\beta_2^{-\frac{r+1}{\beta_2}} \frac{1}{\text{SNR}} & = \left[ \frac{r}{\beta_2^{r+1}} \right]^{\frac{1}{\beta_2}} \left[ \frac{\text{SNR}}{r} \right]^{\frac{1}{\beta_2}} \\
& \leq \left[ \max_r \frac{r}{\beta_2^{r+1}} \right]^{\frac{1}{\beta_2}} \left[ \frac{\text{SNR}}{r} \right]^{\frac{1}{\beta_2}} \\
& \leq \left[ \frac{1}{e\beta_2 \log(\beta_2)} \right]^{\frac{1}{\beta_2}} \left[ \frac{\text{SNR}}{r} \right]^{\frac{1}{\beta_2}} \\
& \leq \left[ \frac{\text{SNR}}{r} \right]^{\frac{1}{\beta_2}} \\
& = \frac{1}{\log\left(\frac{r}{\text{SNR}}\right)} \label{eq:82}
\end{align*}
\]
Equation (82) holds since \(\beta_2 \gg 1\) for \(\text{SNR} \to 0\), which makes
\[
\left[ \frac{1}{e\beta_2 \log(\beta_2)} \right]^{\frac{1}{\beta_2}} \leq [1]^{\frac{1}{\beta_2}} = 1.
\]
Combining (79,80,83), we have
\[
M^*(\text{SNR}) \leq \frac{[\log(\frac{r}{\text{SNR}})]^2 + 1}{\log(\frac{r}{\text{SNR}})}. \quad (84)
\]
Finally, using (75), Theorem 1 and (84), we have
\[
\begin{align*}
\frac{r \text{SNR} - r \text{SNR}}{\log(\frac{r}{\text{SNR}})} [\log \log(\frac{r}{\text{SNR}})]^2 + 1 \log(\frac{r}{\text{SNR}}) + o(\text{SNR}^2) \\
& \leq C_{\text{on-off}}(\text{SNR}) \leq r \text{SNR} - r \text{SNR} \frac{\log \log(\frac{r}{\text{SNR}})}{\log(\frac{r}{\text{SNR}})} + o(\text{SNR}^2).
\end{align*}
\]
Since on-off signaling may not be optimal
\[
C_{\text{on-off}}(\text{SNR}) \leq C(\text{SNR}). \quad (86)
\]
As conditioning reduces entropy, we can express the input-output mutual information as

\[ I(x, y) \leq \sum_{k=1}^{r} I(x, y_k). \] (87)

Each term on the right hand side of (87) is maximized by an on-off distribution [11], and we know from [25] that with this distribution, the mutual information \( \forall k \in \{1 \ldots r\} \) is

\[ I(x, y_k) \leq \text{SNR} - \text{SNR} \frac{\log \log \left( \frac{1}{\text{SNR}} \right)}{\log \left( \frac{\text{SNR}}{r} \right)} + o(\text{SNR}^2). \]

Hence, we can upper bound the capacity as

\[ C(\text{SNR}) \leq r \text{SNR} - r \text{SNR} \frac{\log \log \left( \frac{1}{\text{SNR}} \right)}{\log \left( \frac{\text{SNR}}{r} \right)} + o(\text{SNR}^2). \]

Since,

\[ \frac{\log \log \left( \frac{r}{\text{SNR}} \right)}{\log \left( \frac{1}{\text{SNR}} \right)} \leq \frac{\log \log \left( \frac{1}{\text{SNR}} \right)}{\log \left( \frac{1}{\text{SNR}} \right)}, \]

we have

\[ C(\text{SNR}) \leq r \text{SNR} - r \text{SNR} \frac{\log \log \left( \frac{r}{\text{SNR}} \right)}{\log \left( \frac{1}{\text{SNR}} \right)} + o(\text{SNR}^2). \] (88)

Combining (85, 86, 88), we obtain

\[ r \text{SNR} - r \text{SNR} \frac{\left( \log \log \left( \frac{r}{\text{SNR}} \right) \right)^2 + 1}{\log \left( \frac{1}{\text{SNR}} \right)} + o(\text{SNR}^2) \leq C(\text{SNR}) \leq r \text{SNR} - r \text{SNR} \frac{\log \log \left( \frac{r}{\text{SNR}} \right)}{\log \left( \frac{1}{\text{SNR}} \right)} + o(\text{SNR}^2). \]

We now introduce a notation for the approximation that ignores higher order logarithm functions. Let \( f(\text{SNR}) \) and \( g(\text{SNR}) \) be functions of \( \text{SNR} \). We will denote

\[ f(\text{SNR}) \doteq g(\text{SNR}), \]

if

\[ \lim_{\text{SNR} \to 0} \frac{\log f(\text{SNR})}{\log g(\text{SNR})} = 1. \]

With this scaling, the inequalities in (89) become equalities and the capacity can be expressed as

\[ C(\text{SNR}) = r \text{SNR} - \Delta_{i.i.d}^{(t,r)}(\text{SNR}). \]

where,

\[ \Delta_{i.i.d}^{(t,r)}(\text{SNR}) = \frac{r \text{SNR}}{\log \left( \frac{1}{\text{SNR}} \right)}. \]

Moreover, we also see that on-off signaling (58) is capacity achieving for the i.i.d Rayleigh fading MIMO channel in the wideband regime. (Keeping in mind our scaling, which ignores higher order logarithm functions.)
References


49