On MIMO Fading Channels with Side Information at the Transmitter

EE8510 Project, May 2005
Tairan Wang
twang@ece.umn.edu, ID: 3140947

Abstract

Transmission of information over a MIMO discrete-time flat fading channel with an average power constraint is considered. The scenario in which the transmitter alone has knowledge of the fading levels, as side information, is assumed. This knowledge may be provided in either a causal or a non-causal manner. Upper and lower bounds are derived for each of the two cases. The tools developed are applied to the On/Off fading channel, and some useful strategies for transmitter adaption are discussed.

I. Introduction

Among the extensive research on fading channels, special attention has been given to the scenarios in which the fading coefficients are available to the communication system as side information. These scenarios include the case of channel side information (CSI) available to both the receiver and the transmitter and the case of CSI available to the receiver alone. In the non-coherent scenario, CSI is not available to either the receiver or the transmitter, was considered. The scenario in which the fading levels are available only to the transmitter was left mostly unconsidered. Aside from the scientific curiosity, this scenario gains practical use as well, for example, in OFDM-Discrete Multitone based systems who have a-priori knowledge of all sub-carriers gains. Another motivation comes from the increase in computation resources at cellular base stations which may use the available causal fading levels to design more sophisticated and powerful codes. Such is the case in Time Division Duplexing (TDD) based systems where reciprocity facilitates channel measuring more accurately at the transmitter, due to these increased processing capabilities. The receiver, for complexity reasons, avoids this operation.

This project checks the results in paper [1] carefully and then extend them to a MIMO counterpart. In [1], problem of communicating through a flat fading AWGN channel with the fading coefficients available to the transmitter, as side information, was considered, in either a causal or a non-causal manner. Upper and lower bounds are derived for each of the two cases and Arimoto-Blahut like algorithms, to numerically compute capacity in each case, are presented when an average power constraint was imposed. The tools developed are applied to the On/Off fading channel, and to some restricted cases of a Rayleigh fading channel. In the latter case the capacity per unit cost is examined and it is shown that transmitting at an arbitrarily low E_b/N_0 will sustain reliable communication at zero spectral efficiency, regardless of the causality/non-causality nature of the available side information, mimicking the case of fully available CSI at the transmitter and the receiver.

II. Channel Model

There are M transmitter antennas and N receiver antennas. So, we consider the following MIMO channel model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & \cdots & s_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}, \tag{1}$$

where $x_m \in \mathcal{C}$ is the channel input from the mth transmitter antenna, $y_n \in \mathcal{C}$ is the channel output at the nth receiver antenna, $s_{nm} \in \mathcal{C}$ is the flat fading coefficient from transmitter antenna m to receiver antenna n. The fading coefficients are independent (with respect to both n and m) and identically distributed (i.i.d.). The additive noise at receiver antenna n is denoted $z_n \in \mathcal{C}$, and is independent (with respect to n), identically distributed $\mathcal{CN}(0,1)$. An equivalent channel is

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{z}, \tag{2}$$

where $\mathbf{y} := [y_1, \ldots, y_N]^T$, $\mathbf{x} := [x_1, \ldots, x_M]^T$, $\mathbf{z} := [z_1, \ldots, z_N]^T$, and $\mathbf{S} := \begin{bmatrix} s_{11} & \cdots & s_{1M} \\ \vdots & \ddots & \vdots \\ s_{N1} & \cdots & s_{NM} \end{bmatrix}$. The fading coefficients $\mathbf{S} = \mathbf{S}(i)$ vary with time i (discrete time index). We assume the noise processes are independent of the fading processes and of the channel inputs. We further assume a perfect knowledge of the fading coefficients $\mathbf{S}(i)$ at the transmitter in either a causal manner $\{\mathbf{S}(k)| -\infty \le k \le i\}$ or a non-causal manner $\{\mathbf{S}(k)| -\infty \le k \le \infty\}$. Finally, it is assumed that the signalling is subject to the average power constraint

$$E[\|\mathbf{x}\|^2] \le P. \tag{3}$$

Gel'fand and Pinsker [3] have found the capacity formula for a discrete memoryless channel with random state S known non-causally to the transmitter. Following the extensions made by Costa [5] for discrete-time channels with continuous alphabets and the introduction of constraints on the channel input X we have

$$C_{nc} = \sup_{p(\mathbf{u}|\mathbf{S}),\ \mathcal{F}:\mathcal{U}\times\mathcal{S}\to\mathcal{X},\ E[\|\mathcal{F}(U,S)\|^2]\le P} \left\{ I(U;Y) - I(U;S) \right\}, \tag{4}$$

where U is an auxiliary random variable, $\mathcal{F}$ is a deterministic function, and the joint distribution of the random variables S, U, X and Y is given by

$$p(\mathbf{S}, \mathbf{u}, \mathbf{x}, \mathbf{y}) = \begin{cases} p(\mathbf{S})p(\mathbf{u}|\mathbf{S})p(\mathbf{y}|\mathbf{x}, \mathbf{S}) & \text{if } \mathbf{x} = \mathcal{F}(\mathbf{u}, \mathbf{S}), \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

It was shown by Cohen [4] that taking U to be independent of S leads to a capacity formula equivalent to that given by Shannon [2] corresponding to the problem where the transmitter has causal CSI. In the case we have

$$C_c = \sup_{p(\mathbf{u}),\ \mathcal{F}:\mathcal{U}\times\mathcal{S}\to\mathcal{X},\ E[\|\mathcal{F}(U,S)\|^2]\le P} I(U;Y), \tag{6}$$

where the joint distribution of the random variables S, U, X and Y is given by

$$p(\mathbf{S}, \mathbf{u}, \mathbf{x}, \mathbf{y}) = \begin{cases} p(\mathbf{S})p(\mathbf{u})p(\mathbf{y}|\mathbf{x}, \mathbf{S}) & \text{if } \mathbf{x} = \mathcal{F}(\mathbf{u}, \mathbf{S}), \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Similarly, the capacity formula form given by Shannon (extended to continuous alphabets and including an average power constraint),

$$C_c = \sup_{p(\mathbf{t}),\ E[\|T(S)\|^2]\le P} I(T;Y), \tag{8}$$

can be shown to be a special case of the Gel'fand-Pinsker formula by taking the strategies probability distribution to be independent of the channel state,

$$C_{nc} = \sup_{p(\mathbf{t}|\mathbf{S}),\ E[\|T(S)\|^2]\le P} \{I(T;Y) - I(T;S)\}, \tag{9}$$

where $\mathbf{t} \in \mathcal{T}$, the set of all possible mappings $\mathbf{t} : \mathcal{S} \to \mathcal{X}$ which we will refer to as Shannon strategies or simply as strategies. We will use either form of each capacity formulae as suitable.

III. Lower Bounds

The lower bound can be obtained by choosing an appropriate strategy. For the non-causal case, we examine

$$\begin{aligned} I(U;Y) - I(U;S) &= I(U;S,Y) - I(U;S|Y) - (I(U;S,Y) - I(U;Y|S)) \\ &= I(U;Y|S) - I(U;S|Y). \end{aligned} \tag{10}$$

The lower bound will be obtained by using the following choice of conditional probability distributions:

$$p_L(\mathbf{x}|\mathbf{S}) = \arg\left\{ \sup_{p(\mathbf{x}|\mathbf{S})\in\Omega} I(X;Y|S) \right\} \text{ provided it exists,} \tag{11}$$

$$p_L(\mathbf{u}|\mathbf{x}, \mathbf{S}) = \begin{cases} \delta(\mathbf{u} - g_\beta(\mathbf{x}, \mathbf{S})) & \text{if} \mathbf{S} \in \mathcal{S}, \\ Q_\beta(\mathbf{u}|\mathbf{S}) & \text{if} \mathbf{S} \in \overline{\mathcal{S}}, \end{cases} \tag{12}$$

using the following definitions,

$$
\begin{aligned}
\mathcal{S} &= \{\mathbf{S}|I(X;Y|\mathbf{S}) \neq 0, X \sim p_L(\mathbf{x}|\mathbf{S})\}, \\
\overline{\mathcal{S}} &= \{\mathbf{S}|I(X;Y|\mathbf{S}) = 0, X \sim p_L(\mathbf{x}|\mathbf{S})\}, \\
\beta &\equiv P/N,
\end{aligned}
$$

where g_β is a chosen function which depends on the parameter β (the SNR) and with the requirement that there is a one-to-one mapping between $\mathbf{x}$ and $\mathbf{u}$ for every $\mathbf{S} \in \mathcal{S}$, and finally Q_β is a chosen conditional probability distribution that depends on β as well. The idea behind taking Q_β to be independent of $\mathbf{x}$ is that for those $\mathbf{S}$ which it is defined for X has zero power.

As for the causal case, the lower bounds will be derived from eq. (6) taking U to be a Gaussian random variable and X given $S = \mathbf{S}$ to be Gaussian as well.

IV. Upper Bounds

It is well known that when complete side information is given to both the transmitter and the receiver the capacity of channel (2) is given by [8],

$$
\sup_{p(\mathbf{x}|\mathbf{S}) \in \Omega} I(X;Y|S), \tag{13}
$$

where Ω is the set of all conditional probability distributions $p(\mathbf{x}|\mathbf{S})$ satisfying the constraint (3). Note that eq. (13) may be regarded as a special case of the Gel'fand-Pinsker model when the state S is added to the observation Y at the receiver's end. This, of course, is a trivial upper bound on the channel capacity when side information is available only to the transmitter.

A. Non-Causal CSI

In order to develop an upper bound on Gel'fand-Pinsker capacity, consider the capacity formulation given by eq. (9). Furthermore, for the time being we assume all the relevant alphabets are discrete. The associated average power constraint is seen to be,

$$
\sum_{\mathbf{S},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})\|\mathbf{t}(\mathbf{S})\|^2 \leq P. \tag{14}
$$

Expanding the mutual information we see that

$$
\begin{aligned}
I(T;Y) - I(T;S) &= \sum_{\mathbf{S},\mathbf{y},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S}) \ln \frac{p(\mathbf{t}|\mathbf{y})}{p(\mathbf{t}|\mathbf{S})} \\
&= \sum_{\mathbf{S},\mathbf{y},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S}) \left\{ \ln \frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})} + \ln \frac{q(\mathbf{y})}{p(\mathbf{y})} \right\} \\
&= \sum_{\mathbf{S},\mathbf{y},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S}) \ln \frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})} - D(p(\mathbf{y})\|q(\mathbf{y})) \\
&\leq \sum_{\mathbf{S},\mathbf{y},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S}) \ln \frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})},
\end{aligned}
$$

for any probability distribution $q(\mathbf{y})$ on Y, where equality is achieved iff $q(\mathbf{y}) = p(\mathbf{y})$. Finding the capacity can now be rewritten as the following optimization problem,

$$
C_{nc} = \max_{p(\mathbf{t}|\mathbf{S})} \min_{q(\mathbf{y})} \sum_{\mathbf{S},\mathbf{y},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S}) \ln \frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}, \tag{15}
$$

with the constraints, $\sum_{\mathbf{t}} p(\mathbf{t}|\mathbf{S}) = 1 \ \forall \mathbf{S}$, $p(\mathbf{t}|\mathbf{S}) \geq 0 \ \forall \mathbf{S}, \mathbf{t}$, $\sum_{\mathbf{S},\mathbf{t}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})\|\mathbf{t}(\mathbf{S})\|^2 \leq P$.

Lemma 1: The functional over which we optimize in (15) is concave in $p(\mathbf{t}|\mathbf{S})$ and convex in $q(\mathbf{y})$.

Proof is similar to the counterpart in paper [1] and is omitted here. Following the work in [7] we can reformulate our maximization problem using the lagrange dual technique. Forming a partial lagrangian for (15) we have,

$$\mathcal{L}=\sum_{\mathbf{S},\mathbf{t},\mathbf{y}} p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}+\sum_{\mathbf{S}}\lambda_{1\mathbf{S}}\left(1-\sum_{\mathbf{t}}p(\mathbf{t}|\mathbf{S})\right)+\lambda_2\left(P-\sum_{\mathbf{S},\mathbf{t}}p(\mathbf{S})p(\mathbf{t}|\mathbf{S})\|\mathbf{t}(\mathbf{S})\|^2\right), \quad (16)$$

where we introduce the lagrange multipliers $\lambda_{1\mathbf{S}}\in\mathcal{R},\ \lambda_2\geq 0$ and notice that $\max_{p(\mathbf{t}|\mathbf{S})}\min_{q(\mathbf{y})}\min_{\lambda_{1\mathbf{S}},\lambda_2}\mathcal{L}$, such that $p(\mathbf{t}|\mathbf{S})\geq 0$ and $q(\mathbf{y})$ is a valid probability distribution, has the same optimal value as (15). Continuing to follow the ideas in [7] (note the exchange of the $\min$ and the $\max$) we have the upper bound on problem (15) given by

$$C_{nc}\leq\min_{q(\mathbf{y}),\lambda_{1\mathbf{S}},\lambda_2}\max_{p(\mathbf{t}|\mathbf{S})}\left\{\sum_{\mathbf{S},\mathbf{t}}p(\mathbf{S})p(\mathbf{t}|\mathbf{S})\times\left[\sum_{\mathbf{y}}p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{p(\mathbf{y},\mathbf{t})}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\frac{\lambda_{1\mathbf{S}}}{p(\mathbf{S})}-\lambda_2\|\mathbf{t}(\mathbf{S})\|^2\right]+\sum_{\mathbf{S}}\lambda_{1\mathbf{S}}+\lambda_2 P\right\}, \quad (17)$$

where $p(\mathbf{t}|\mathbf{S})\geq 0$ and $q(\mathbf{y})$ is a valid probability distribution. To find the solving $p(\mathbf{t}|\mathbf{S})$ of the inner maximization, for a given $q(\mathbf{y})$, $\lambda_{1\mathbf{S}}$ and λ_2, we differentiate $\mathcal{L}$ with respect to $p(\mathbf{t}|\mathbf{S})$,

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial p(\mathbf{t}|\mathbf{S})} &= \sum_{\mathbf{y}}p(\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{y}|\mathbf{t},\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\lambda_{1\mathbf{S}}-p(\mathbf{S})\lambda_2\|\mathbf{t}(\mathbf{S})\|^2\\ &\quad+\sum_{\mathbf{y}}p(\mathbf{S})p(\mathbf{t}|\mathbf{S})p(\mathbf{y}|\mathbf{t},\mathbf{S})\frac{-1}{p(\mathbf{t}|\mathbf{S})}+\sum_{\mathbf{S}',\mathbf{y}}p(\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{y}|\mathbf{t},\mathbf{S}')\frac{p(\mathbf{y}|\mathbf{t},\mathbf{S})p(\mathbf{S})}{\sum_{\mathbf{S}''}p(\mathbf{S}'')p(\mathbf{t}|\mathbf{S}'')p(\mathbf{y}|\mathbf{t},\mathbf{S}'')}\\ &= p(\mathbf{S})\left\{\sum_{\mathbf{y}}p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{y}|\mathbf{t},\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\frac{\lambda_{1\mathbf{S}}}{p(\mathbf{S})}-\lambda_2\|\mathbf{t}(\mathbf{S})\|^2\right\}\end{aligned}$$

and we arrive to the following conditions on the solution to the maximization problem (17) (applying the Karush-Kuhn-Tucker (KKT) conditions)

$$\begin{aligned}&\sum_{\mathbf{y}}p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{y}|\mathbf{t},\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\frac{\lambda_{1\mathbf{S}}}{p(\mathbf{S})}-\lambda_2\|\mathbf{t}(\mathbf{S})\|^2=0,\ \text{if } p(\mathbf{t}|\mathbf{S})>0,\\ &\sum_{\mathbf{y}}p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{y}|\mathbf{t},\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\frac{\lambda_{1\mathbf{S}}}{p(\mathbf{S})}-\lambda_2\|\mathbf{t}(\mathbf{S})\|^2\leq 0,\ \text{if } p(\mathbf{t}|\mathbf{S})=0.\end{aligned} \quad (18)$$

With the above conditions we can see that for the maximizing $p(\mathbf{t}|\mathbf{S})$ in (17) we have

$$\min_{q(\mathbf{y}),\lambda_{1\mathbf{S}},\lambda_2}\max_{p(\mathbf{t}|\mathbf{S})}\mathcal{L}=\min_{q(\mathbf{y}),\lambda_{1\mathbf{S}},\lambda_2}\sum_{\mathbf{S}}\lambda_{1\mathbf{S}}+\lambda_2 P. \quad (19)$$

To conclude, we now have a dual problem to the problem (17) (the solution of which is an upper bound on problem (15)):

$$C_{nc}(P)\leq\min_{q(\mathbf{y}),\lambda_{1\mathbf{S}},\lambda_2}\sum_{\mathbf{S}}\lambda_{1\mathbf{S}}+\lambda_2 P, \quad (20)$$

where the set over which we minimize consists of valid probability distributions $q(\cdot)$ on Y, real numbers $\lambda_{1\mathbf{S}}$ and $\lambda_2\geq 0$ for which a function $p(\mathbf{t}|\mathbf{S})\geq 0$ exists such that,

$$\sum_{\mathbf{y}}p(\mathbf{y}|\mathbf{t},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{y}|\mathbf{t},\mathbf{S}')p(\mathbf{t}|\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})p(\mathbf{t}|\mathbf{S})}-\frac{\lambda_{1\mathbf{S}}}{p(\mathbf{S})}-\lambda_2\|\mathbf{t}(\mathbf{S})\|^2\leq 0,\quad\forall\ \mathbf{S},\mathbf{t}, \quad (21)$$

and for any $\mathbf{S},\mathbf{t}$ such that $p(\mathbf{t}|\mathbf{S})\geq 0$ equality in eq. (21) must hold. Choosing any feasible $\lambda_{1\mathbf{S}},\ \lambda_2$ and $q(\mathbf{y})$ in eq. (20) such that (21) holds true will give us an upper bound on the Gel'fand-Pinsker capacity with an average power constraint.

B. Causal CSI

Again starting with the assumption that all relevant alphabets are discrete and examining eq. (6) we have the following,

Lemma 2: For every probability distribution $p(\cdot)$ on U, deterministic function $\mathcal{F}(\mathbf{u},\mathbf{S})$ and probability distribution $q(\cdot)$ on the channel output Y

$$I(U;Y) \leq \sum_{\mathbf{u}} p(\mathbf{u}) D(w_{\mathcal{F}}(\cdot|\mathbf{u}) \| q(\cdot)), \tag{22}$$

where

$$D(w_{\mathcal{F}}(\cdot|\mathbf{u})\|q(\cdot)) = \sum_{\mathbf{y}} \sum_{\mathbf{S}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S}) p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}'),\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})}.$$

Proof: we know that,

$$\begin{aligned} I(U;Y) &= \sum_{\mathbf{S},\mathbf{u},\mathbf{x},\mathbf{y}} p(\mathbf{y}|\mathbf{x},\mathbf{S})p(\mathbf{x}|\mathbf{u},\mathbf{S})p(\mathbf{u})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}',\mathbf{x}'} p(\mathbf{y}|\mathbf{x}',\mathbf{S}')p(\mathbf{x}'|\mathbf{u},\mathbf{S}')p(\mathbf{u})p(\mathbf{S}')}{p(\mathbf{u})\sum_{\mathbf{S}'',\mathbf{u}'',\mathbf{x}''} p(\mathbf{y}|\mathbf{x}'',\mathbf{S}'')p(\mathbf{x}''|\mathbf{u}'',\mathbf{S}'')p(\mathbf{u}'')p(\mathbf{S}'')} \\ &= \sum_{\mathbf{S},\mathbf{u},\mathbf{y}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S})p(\mathbf{u})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}'),\mathbf{S}')p(\mathbf{S}')}{\sum_{\mathbf{S}'',\mathbf{u}''} p(\mathbf{y}|\mathcal{F}(\mathbf{u}'',\mathbf{S}''),\mathbf{S}'')p(\mathbf{u}'')p(\mathbf{S}'')}, \end{aligned}$$

now for every $p(\mathbf{u})$, $\mathcal{F}(\mathbf{u},\mathbf{S})$ and $q(\mathbf{y})$ we have,

$$\begin{aligned} &\sum_{\mathbf{u}} p(\mathbf{u}) \sum_{\mathbf{y},\mathbf{S}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}'),\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})} - I(U;Y) \\ &= \sum_{\mathbf{u}} p(\mathbf{u}) \sum_{\mathbf{y},\mathbf{S}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'',\mathbf{u}''} p(\mathbf{y}|\mathcal{F}(\mathbf{u}'',\mathbf{S}''),\mathbf{S}'')p(\mathbf{S}'')p(\mathbf{u}'')}{q(\mathbf{y})} \\ &= \sum_{\mathbf{y}} \sum_{\mathbf{S},\mathbf{u}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S})p(\mathbf{u})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'',\mathbf{u}''} p(\mathbf{y}|\mathcal{F}(\mathbf{u}'',\mathbf{S}''),\mathbf{S}'')p(\mathbf{S}'')p(\mathbf{u}'')}{q(\mathbf{y})} \\ &= D(\tilde{p}(\mathbf{y})\|q(\mathbf{y})) \geq 0, \end{aligned}$$

therefore,

$$\begin{aligned} I(U;Y) &\leq \sum_{\mathbf{u}} p(\mathbf{u}) \sum_{\mathbf{y}} \sum_{\mathbf{S}} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}),\mathbf{S})p(\mathbf{S}) \ln \frac{\sum_{\mathbf{S}'} p(\mathbf{y}|\mathcal{F}(\mathbf{u},\mathbf{S}'),\mathbf{S}')p(\mathbf{S}')}{q(\mathbf{y})} \\ &= \sum_{\mathbf{u}} p(\mathbf{u}) D(w_{\mathcal{F}}(\cdot|\mathbf{u})\|q(\cdot)). \end{aligned}$$

Note we write $w_{\mathcal{F}}(\cdot|\cdot)$ to emphasize the dependence of w on $\mathcal{F}$. The extension of the upper bound (22) to continuous alphabets and to constrained input may be done to obtain:

$$\begin{aligned} I(U;Y) &\leq \int D(W_{\mathcal{F}}(\cdot|\mathbf{u})\|Q(\cdot))dP(\mathbf{u}) \\ &\leq \sup_{\mathbf{u}} D(W_{\mathcal{F}}(\cdot|\mathbf{u})\|Q(\cdot)), \end{aligned} \tag{23}$$

and the upper bound:

$$\begin{aligned} C_c(P) &= \sup_{P(\mathbf{u}),\mathcal{F}} \inf_{\gamma\geq 0} \left\{ I(U;Y) + \gamma\left(P - \iint \|\mathcal{F}(\mathbf{u},\mathbf{S})\|^2 dP(\mathbf{S})dP(\mathbf{u})\right)\right\} \\ &\leq \inf_{\gamma\geq 0} \sup_{P(\mathbf{u}),\mathcal{F}} \left\{ I(U;Y) + \gamma\left(P - \iint \|\mathcal{F}(\mathbf{u},\mathbf{S})\|^2 dP(\mathbf{S})dP(\mathbf{u})\right)\right\} \\ &\leq \inf_{\gamma\geq 0} \sup_{P(\mathbf{u}),\mathcal{F}} \sup_{\mathbf{u}} \left\{ D(W_{\mathcal{F}}(\cdot|\mathbf{u})\|Q(\cdot)) + \gamma\left(P - \int \|\mathcal{F}(\mathbf{u},\mathbf{S})\|^2 dP(\mathbf{S})\right)\right\} \\ &= \inf_{\gamma\geq 0} \sup_{\mathcal{F}} \sup_{\mathbf{u}} \left\{ D(W_{\mathcal{F}}(\cdot|\mathbf{u})\|Q(\cdot)) + \gamma\left(P - \int \|\mathcal{F}(\mathbf{u},\mathbf{S})\|^2 dP(\mathbf{S})\right)\right\}. \end{aligned} \tag{24}$$

V. Applications and Discussion

To give an application example, we consider parallel fading channel, which is a degraded case of MIMO channel. In parallel fading channel, the number of transmitter antennas is the same as the number of receiver antennas, i.e., $M = N = K$; and the channel $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_K)$ is a diagonal matrix.

More important, by considering parallel fading channel, we can set the channel in real field $\mathcal{R}$ instead of complex field $\mathcal{C}$, because after using any strategy and before transmitting, we can always add an extra phase to the transmitted signal to cancel the phase of the corresponding channel coefficient while satisfying the average power constraint (at least making $s_1, \ldots, s_K$ to be real). Then, at the receiver side, it can always separate the real part and imaginary part of the received signals and process them respectively. And the two parts have the same statistics, so analyze one part is enough to understand the whole story. So the following analysis will be in the real field $\mathcal{R}$. We will take $K = 2$ in the following example.

A. *On/Off Fading Channel*

In this example we shall consider a channel with binary fading, that is: $P_r(s_1 = 1) = 1 - P_r(s_1 = 0) = \alpha_1$, $P_r(s_2 = 1) = 1 - P_r(s_2 = 0) = \alpha_2$. To find the lower bound on the non-causal CSI capacity we start by solving eq. (11) which gives (note the average power constraint (3) holds):

$$f_L(\mathbf{x}|\mathbf{S}) = \begin{cases} \delta(x_1)\delta(x_2) & s_1 = 0, s_2 = 0 \\ \frac{1}{\sqrt{2\pi P_1}} \exp(-\frac{x_1^2}{2P_1})\delta(x_2) & s_1 = 1, s_2 = 0 \\ \frac{1}{\sqrt{2\pi P_2}} \exp(-\frac{x_2^2}{2P_2})\delta(x_1) & s_1 = 0, s_2 = 1 \\ \frac{1}{\pi P_3} \exp(-\frac{x_1^2+x_2^2}{P_3}) & s_1 = 1, s_2 = 1 \end{cases} \tag{25}$$

where $P_1 = P_2 = \frac{P}{\alpha_1+\alpha_2}$ and $P_3 = \frac{2P}{\alpha_1+\alpha_2}$ by waterfilling over both space and time. For eq. (12) we shall take $g_\beta(\mathbf{x}, \mathbf{S}) = g_{1\beta}(x_1, s_1)g_{2\beta}(x_2, s_2)$, $g_{1\beta}(x_1, s_1 = 1) = x_1$, $g_{2\beta}(x_2, s_2 = 1) = x_2$, and $Q_\beta(\mathbf{u}|\mathbf{S}) = Q_{1\beta}(u_1|s_1)Q_{2\beta}(u_2|s_2)$, $Q_{1\beta}(u_1|s_1 = 0) = \mathcal{N}(0, N{\psi_1}^2)$, $Q_{2\beta}(u_2|s_2 = 0) = \mathcal{N}(0, N{\psi_2}^2)$. Thus, we have

$$f_L(\mathbf{u}|\mathbf{x}, \mathbf{S}) = f_L(u_1|x_1, s_1)f_L(u_2|x_2, s_2) \tag{26}$$

$$f_L(u_1|x_1, s_1) = \begin{cases} \delta(u_1 - x_1) & s_1 = 1 \\ \frac{1}{\sqrt{2\pi N{\psi_1}^2}} \exp(-\frac{u_1^2}{2N{\psi_1}^2}) & s_1 = 0 \end{cases} \tag{27}$$

$$f_L(u_2|x_2, s_2) = \begin{cases} \delta(u_2 - x_2) & s_2 = 1 \\ \frac{1}{\sqrt{2\pi N{\psi_2}^2}} \exp(-\frac{u_2^2}{2N{\psi_2}^2}) & s_2 = 0 \end{cases} \tag{28}$$

where ψ_1 and ψ_2 depend on P/N and will be determined through numerical optimization. Using eq. (25)$\sim$ (28) and after some work we get:

$$\begin{aligned} p(\mathbf{u}|\mathbf{S} = \mathrm{diag}(0,0)) &= \frac{1}{\sqrt{2\pi N\psi_1^2}} \exp(-\frac{u_1^2}{2N\psi_1^2}) \frac{1}{\sqrt{2\pi N\psi_2^2}} \exp(-\frac{u_2^2}{2N\psi_2^2}) \\ p(\mathbf{u}|\mathbf{S} = \mathrm{diag}(1,0)) &= \frac{1}{\sqrt{2\pi P/(\alpha_1+\alpha_2)}} \exp(-\frac{u_1^2}{2P/(\alpha_1+\alpha_2)}) \frac{1}{\sqrt{2\pi N\psi_2^2}} \exp(-\frac{u_2^2}{2N\psi_2^2}) \\ p(\mathbf{u}|\mathbf{S} = \mathrm{diag}(0,1)) &= \frac{1}{\sqrt{2\pi P/(\alpha_1+\alpha_2)}} \exp(-\frac{u_2^2}{2P/(\alpha_1+\alpha_2)}) \frac{1}{\sqrt{2\pi N\psi_1^2}} \exp(-\frac{u_1^2}{2N\psi_1^2}) \\ p(\mathbf{u}|\mathbf{S} = \mathrm{diag}(1,1)) &= \frac{1}{2\pi P/(\alpha_1+\alpha_2)} \exp(-\frac{u_1^2}{2P/(\alpha_1+\alpha_2)}) \exp(-\frac{u_2^2}{2P/(\alpha_1+\alpha_2)}), \end{aligned} \tag{29}$$

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{u},\mathbf{S}=\mathrm{diag}(0,0)) &= \frac{1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N}) \\
p(\mathbf{y}|\mathbf{u},\mathbf{S}=\mathrm{diag}(1,0)) &= \frac{1}{\sqrt{2\pi N}}\exp(-\frac{(y_1-u_1)^2}{2N})\frac{1}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N}) \\
p(\mathbf{y}|\mathbf{u},\mathbf{S}=\mathrm{diag}(0,1)) &= \frac{1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1}{\sqrt{2\pi N}}\exp(-\frac{(y_2-u_2)^2}{2N}) \\
p(\mathbf{y}|\mathbf{u},\mathbf{S}=\mathrm{diag}(0,0)) &= \frac{1}{\sqrt{2\pi N}}\exp(-\frac{(y_1-u_1)^2}{2N})\frac{1}{\sqrt{2\pi N}}\exp(-\frac{(y_2-u_2)^2}{2N}). \qquad (30)
\end{aligned}
$$

Using $p(\mathbf{S}=\mathrm{diag}(0,0))=(1-\alpha_1)(1-\alpha_2)$, $p(\mathbf{S}=\mathrm{diag}(1,0))=\alpha_1(1-\alpha_2)$, $p(\mathbf{S}=\mathrm{diag}(0,1))=(1-\alpha_1)\alpha_2$, $p(\mathbf{S}=\mathrm{diag}(1,1))=\alpha_1\alpha_2$, and eq. (29) (30) in eq. (10), finally we reach the lower bound eq. (31) (no space for closed form).

$$
\begin{aligned}
C_{nc} &\geq I(U;Y)-I(U;S) \\
&= \iint \sum_{\mathbf{S}} p(\mathbf{S})p(\mathbf{u}|\mathbf{S})p(\mathbf{y}|\mathbf{u},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{S}')p(\mathbf{u}|\mathbf{S}')p(\mathbf{y}|\mathbf{u},\mathbf{S}')}{\int\sum_{\mathbf{S}''}p(\mathbf{S}'')p(\mathbf{u}'|\mathbf{S}'')p(\mathbf{y}|\mathbf{u}',\mathbf{S}'')d\mathbf{u}'p(\mathbf{u}|\mathbf{S})}d\mathbf{y}d\mathbf{u}. \qquad (31)
\end{aligned}
$$

The lower bound on the causal CSI capacity is obtained by taking the following probability distribution

$$
\begin{aligned}
p(\mathbf{u}) &= f_L(\mathbf{u})=f_L(u_1)f_L(u_2) \\
f_L(u_1) &= \frac{1}{\sqrt{2\pi P/(\alpha_1+\alpha_2)}}\exp(-\frac{u_1^2}{2P/(\alpha_1+\alpha_2)}) \\
f_L(u_2) &= \frac{1}{\sqrt{2\pi P/(\alpha_1+\alpha_2)}}\exp(-\frac{u_2^2}{2P/(\alpha_1+\alpha_2)})
\end{aligned}
$$

and the deterministic function $\mathcal{F}(\mathbf{u},\mathbf{S})=\mathbf{S}\mathbf{u}$ and from eq. (6) we have the bound eq. (32) (no space for closed form).

$$
\begin{aligned}
C_c &\geq I(U;Y) \\
&= \iint \sum_{\mathbf{S}} p(\mathbf{S})p(\mathbf{u})p(\mathbf{y}|\mathbf{u},\mathbf{S})\ln\frac{\sum_{\mathbf{S}'}p(\mathbf{S}')p(\mathbf{y}|\mathbf{u},\mathbf{S}')}{\int\sum_{\mathbf{S}''}p(\mathbf{S}'')p(\mathbf{u}')p(\mathbf{y}|\mathbf{u}',\mathbf{S}'')d\mathbf{u}'}d\mathbf{y}d\mathbf{u}. \qquad (32)
\end{aligned}
$$

To develop the upper bound on the causal CSI capacity we start by writing eq. (24) for this case,

$$
\begin{aligned}
C_c \leq \inf_{\gamma\geq 0}\sup_{\mathcal{F}}\sup_{\mathbf{u}} \Bigg\{ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[p(\mathbf{y},\mathcal{F},\mathbf{u})]\ln\frac{p(\mathbf{y},\mathcal{F},\mathbf{u})}{q(\mathbf{y})}dy_1dy_2 \\
&+\gamma(P-(1-\alpha_1)(1-\alpha_2)\|\mathcal{F}(\mathbf{u},s_1=0,s_2=0)\|^2-(1-\alpha_1)\alpha_2\|\mathcal{F}(\mathbf{u},s_1=0,s_2=1)\|^2 \\
&-\alpha_1(1-\alpha_2)\|\mathcal{F}(\mathbf{u},s_1=1,s_2=0)\|^2-\alpha_1\alpha_2\|\mathcal{F}(\mathbf{u},s_1=1,s_2=1)\|^2)\Bigg\}, \qquad (33)
\end{aligned}
$$

where $\mathcal{F}(\mathbf{u},\mathbf{S})=[\mathcal{F}(u_1,s_1)\ \mathcal{F}(u_2,s_2)]^T$,

$$
\begin{aligned}
p(\mathbf{y},\mathcal{F},\mathbf{u}) = &\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N}) \\
&+\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{(y_2-\mathcal{F}(u_2,s_2=1))^2}{2N}) \\
&+\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{(y_1-\mathcal{F}(u_1,s_1=1))^2}{2N}) \\
&+\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{(y_1-\mathcal{F}(u_1,s_1=1))^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{(y_2-\mathcal{F}(u_2,s_2=1))^2}{2N}). \qquad (34)
\end{aligned}
$$

It is evident from the equation above that the optimal choice for $\mathcal{F}(u_i, s_i = 0)$ is 0, $i = 1, 2$. Taking

$$\begin{aligned}
q(\mathbf{y}) \;=\;& \frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N}) \\
&+\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{\alpha_2}{\sqrt{2\pi(N+P/(\alpha_1+\alpha_2))}}\exp(-\frac{y_2^2}{2(N+P/(\alpha_1+\alpha_2))}) \\
&+\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\frac{\alpha_1}{\sqrt{2\pi(N+P/(\alpha_1+\alpha_2))}}\exp(-\frac{y_1^2}{2(N+P/(\alpha_1+\alpha_2))}) \\
&+\frac{\alpha_1\alpha_2}{2\pi(N+P/(\alpha_1+\alpha_2))}\exp(-\frac{y_1^2}{2(N+P/(\alpha_1+\alpha_2))})\exp(-\frac{y_2^2}{2(N+P/(\alpha_1+\alpha_2))}),
\end{aligned} \tag{35}$$

and making the assignments $\gamma' = \gamma P$ and $\delta_1 = \sqrt{(\alpha_1+\alpha_2)/P}\mathcal{F}(u_1, s_1 = 1)$, $\delta_2 = \sqrt{(\alpha_1+\alpha_2)/P}\mathcal{F}(u_2, s_2 = 1)$ in eq. (33) we get eq. (36), where we let $\delta_1, \delta_2 \in \mathcal{R}$ and the change of variables allows us to exchange the $\sup$ over $\mathcal{F}$ and $\mathbf{u}$ with the $\sup$ over δ_1 and δ_2.

$$\begin{aligned}
C_c \;\leq\;& \inf_{\gamma'\geq 0}\sup_{\delta_1}\sup_{\delta_2}\Big\{\gamma'(1-\frac{\alpha_1\delta_1^2+\alpha_2\delta_2^2}{\alpha_1+\alpha_2}) \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\ln\frac{p(y_1,y_2,\delta_1,\delta_2)}{q(y_1,y_2)}dy_1dy_2 \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2'^2}{2N})\ln\frac{p(y_1,y_2',\delta_1,\delta_2)}{q(y_1,y_2')}dy_1dy_2' \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1'^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\ln\frac{p(y_1',y_2,\delta_1,\delta_2)}{q(y_1',y_2)}dy_1'dy_2 \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1'^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2'^2}{2N})\ln\frac{p(y_1',y_2',\delta_1,\delta_2)}{q(y_1',y_2')}dy_1'dy_2'\Big\},
\end{aligned} \tag{36}$$

where $y_1' = y_1 - \delta_1\sqrt{\frac{P}{N(\alpha_1+\alpha_2)}}$ and $y_2' = y_2 - \delta_2\sqrt{\frac{P}{N(\alpha_1+\alpha_2)}}$.

Looking at the limit as $\delta_1 \to \infty$ and $\delta_2 \to \infty$ of eq. (36) (similar to [1]), we get a necessary condition for the supremum in (36) to exist:

$$\gamma' \geq \frac{\alpha_1+\alpha_2}{2(1+N(\alpha_1+\alpha_2)/P)}. \tag{37}$$

Taking γ' in (36) to be the right side of eq. (37) we have the upper bound for this case, eq. (38) which will be evaluated numerically.

$$\begin{aligned}
C_c \;\leq\;& \sup_{\delta_1}\sup_{\delta_2}\Big\{\frac{\alpha_1(1-\delta_1^2)+\alpha_2(1-\delta_2^2)}{2(1+N(\alpha_1+\alpha_2)/P)} \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\ln\frac{p(y_1,y_2,\delta_1,\delta_2)}{q(y_1,y_2)}dy_1dy_2 \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1-\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2'^2}{2N})\ln\frac{p(y_1,y_2',\delta_1,\delta_2)}{q(y_1,y_2')}dy_1dy_2' \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1'^2}{2N})\frac{1-\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2^2}{2N})\ln\frac{p(y_1',y_2,\delta_1,\delta_2)}{q(y_1',y_2)}dy_1'dy_2 \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\alpha_1}{\sqrt{2\pi N}}\exp(-\frac{y_1'^2}{2N})\frac{\alpha_2}{\sqrt{2\pi N}}\exp(-\frac{y_2'^2}{2N})\ln\frac{p(y_1',y_2',\delta_1,\delta_2)}{q(y_1',y_2')}dy_1'dy_2'\Big\}.
\end{aligned} \tag{38}$$

The upper bound for non-causal case considered here is

$$C_{nc} \leq C_{TRSI} = \frac{\alpha_1+\alpha_2}{2}\ln(1+\frac{P}{N(\alpha_1+\alpha_2)}), \tag{39}$$

where C_{TRSI} stands for the case where both the transmitter and the receiver have CSI. This is a trivial upper bound on the capacities when CSI is available to the transmitter alone, but it is tighter than the upper bound derived in Subsection IV-A.

Fig. 1 and Fig. 2, in the following pages, display the bounds developed above, $C_{lb-c}, C_{ub-c}, C_{lb-nc}$ (with numerical optimization performed when needed), where $lb-c$ stands for the lower bound on causal CSI capacity, etc', and the capacity C_{TRSI} for several interesting values of α_1 and α_2.

Looking at the figures we note the following:

1) As α_1 and α_2 grow larger the bounds become tighter.
2) There is a clear advantage in knowing the side information in a non-causal manner over causal only (at mid and high SNR levels the upper bound on the causal CSI capacity lies beneath the lower bound on the non-causal CSI capacity).
3) For very large SNR the lower bound on the non-causal CSI capacity becomes tight to the capacity in the fully informed case.

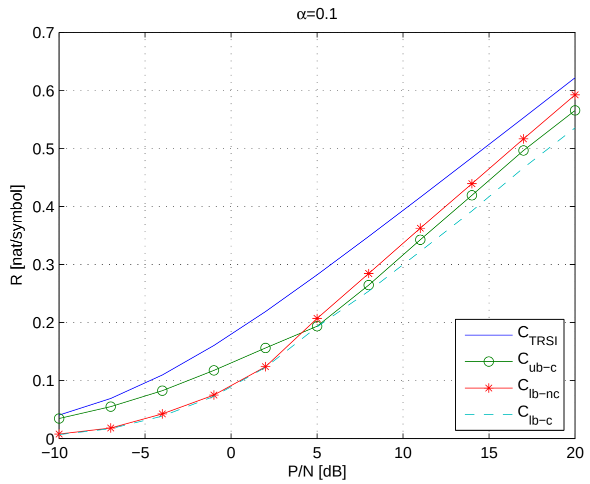

Fig. 1. Bounds on capacity of a parallel binary fading channel with CSI available at the transmitter, $P_r(s_1 = 1) = P_r(s_2 = 1) = 0.1$.

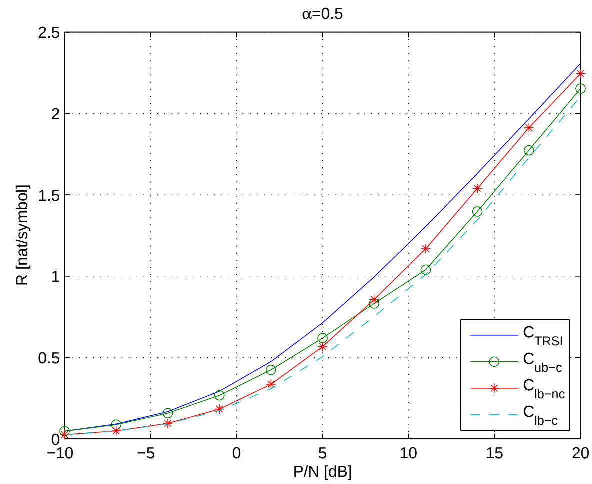

Fig. 2. Bounds on capacity of a parallel binary fading channel with CSI available at the transmitter, $P_r(s_1 = 1) = P_r(s_2 = 1) = 0.5$.

VI. Some Useful Strategies

The most difficult part during the process of finding an optimal strategy is to prove this strategy is optimal. For additive channel in Costa [5], optimal strategy as dirty paper coding can achieve the upper bound, i.e., capacity when complete side information is given to both the transmitter and the receiver. But for fading channel, this upper bound seems to be not tight in general, or say there may be a strict gap between the upper bound and the actual capacity. So, finding a useful (tight) upper bound will be helpful to find the optimal strategy in general. Unfortunately, in paper [1], the upper bound is not tight, especially for non-causal CSI. So we did not consider it in Section V.

A. Channel Inversion

A suboptimal but simple transmitter adaption scheme is channel inversion [9], i.e., it inverts the channel fading. The channel then appears to the encoder and the decoder as a time-invariant AWGN channel. Now, let γ be the instantaneous SNR, and $P(\gamma)$ be the instantaneous transmit power. We have the constraint $\int_\gamma P(\gamma)p(\gamma)d\gamma \leq P$. Then the power adaptation for channel inversion is given by $P(\gamma)/P = \sigma/\gamma$, where σ equals the constant received SNR which can be maintained under the transmit power constraint. The constant σ thus satisfies $\int \sigma/\gamma p(\gamma) = 1$, so $\sigma = 1/E[1/\gamma]$.

The fading channel capacity with channel inversion is just the capacity of an AWGN channel the SNR σ:

$$C(P) = B\log[1+\sigma] = B\log[1+\frac{1}{E[1/\gamma]}]. \tag{40}$$

Channel inversion can exhibit a large capacity penalty in extreme fading environments. For example, in Rayleigh fading $E[1/\gamma]$ is infinite, and thus the capacity with channel inversion is zero (so is the case for on/off channel).

A truncated inversion policy is also considered in [9], but this policy is not applicable for on/off channel, as it requires the receiver to know the threshold, which is equivalent to know the whole channel in this case.

B. *Log-DPC*

We find a useful strategy for scalar discrete-time flat fading channel model

$$y_i = s_i x_i + z_i, \tag{41}$$

where $x_i \geq 0$ (the reason will be shown later) is the channel input, $y_i \in \mathcal{R}$ is the channel output, $s_i \geq 0$ are i.i.d. equivalent real random variables describing the fading coefficients after phase cancellation and z_i are the i.i.d. Gaussian noise samples with variance N. We assume the noise processes are independent of the fading processes and of the channel inputs. We further assume a perfect knowledge of the fading coefficients s_i at the transmitter in a non-causal manner $\{s_k| -\infty \leq k \leq \infty\}$. Finally, it is assumed that the signalling is subject to the average power constraint

$$E[X_i^2] \leq P. \tag{42}$$

Notice when $P/N \to \infty$, we can always take logarithm to both sides of eq. (41) to get a channel with additive interference:

$$\ln y_i = \ln s_i + \ln x_i. \tag{43}$$

Following the way of dirty paper coding (e.g. modulo scheme [6] [10]), we can use this simple strategy: Suppose U is the desired signal (between -1 and 1), $X = \ln x_i$ is the transmitted signal, and $S = \ln s_i$ is the additive interference (known at TX through s_i, but not at RX). Take modulo $[-1, 1]$ operation to get $X = [U - S]_{[-1,1]}$, and what is transmitted here is $x_i = \exp(X)$ (this is the reason for $x_i \geq 0$). Receiver takes logarithm to the received signal $Y = \ln y_i$, then performs modulo $[-1, 1]$ operation to get $Y' = [Y]_{[-1,1]} = [X + S] = [(U - S) + S] = [U]$. So we can claim that for scalar discrete-time flat fading channel, as SNR goes to infinity, C_{nc} will approach the capacity when CSI is known at both transmitter and receiver. This can be shown somehow in paper [1], where at high SNR, even the lower bound of the non-causal case will be close to the TRSI bound.

VII. Conclusion

In this project, we have tried to extend the results in paper [1] to MIMO case and discussed some useful strategies. During this work, we realized that the complexity for the MIMO counterpart increases exponentially with respect to the number of independent channels. The parallel fading channel was taken for simplicity. In my opinion, this work will not attract people's eyes until some simple but meaningful strategies found to be able to easily implement in industry.

References

[1] Dan Goldsmith, Shlomo Shamai (Shitz), and Yossef Steinberg, "On Fading Channels with Side Information at the Transmitter," submitted to *IEEE Transactions on Information Theory*, Nov. 2004.

[2] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289 - 293, Oct. 1958.

[3] S. I. Gel'fand and M. S. Pinsker, "Coding For Channel With Random Parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19 - 31, 1980.

[4] A. S. Cohen, "Communication with Side Information," Graduate Seminar 6.962, MIT, Cambridge, MA, Spring 2001 (http://web.mit.edu/6.962/www/www_ spring_ 2001/schedule.html).

[5] Max H. M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439 - 441, May 1983.

[6] U. Erez and R. Zamir, "Noise prediction for channels with side information at the transmitter," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1610 - 1617, July 2000.

[7] Sriram Vishwanath and Andrea Goldsmith, "A Duality Theory for Channel Capacity," *Proc. 41st Annual Allerton Conference on Communications, Control and Computing*, Allerton, IL, Oct. 2002.

[8] T. Cover and M. Chiang, "Duality Between Channel Capacity and Rate Distortion With Two-Sided State Information," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1629 - 1638, June 2002.

[9] Andrea Goldsmith and Pravin P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986 - 1992, Nov. 1997.

[10] "Multiuser Information Theory", pp. 6, Lecture Notes 19-20.