The effect on capacity computation of decoupling slow fades from fast fades for channels with asymmetric channel side information

Muriel Médard and R. Srikant

Abstract

Generally, slow fades and fast fades in wireless time-varying channels are modeled as being decoupled. To avoid the expense of providing, through feedback, detailed channel side information to the sender, the receiver offers the sender only a coarse representation of the state of the channel. Specifically, the receiver tracks the fast and slow variations of the channel while the sender receives feedback only about the slow variations. We compute an approximate channel capacity in the following sense: each rate smaller than the “approximate” capacity can be achieved for sufficiently large separation between the slow and fast time scales. The difference between the true capacity and the approximate capacity is \( O(\epsilon \log(\epsilon) \log(-\epsilon \log(\epsilon))) \), where \( \epsilon \) is the ratio between the speed of variation of the channel in the macro and micro-states.

1. Model and Main Result

Consider a discrete-time Markovian fading process defined by the stochastic matrix \( A + \epsilon B \), where \( A \) is block-diagonal with \( M \) blocks and the \( i^{th} \) block (which is also a stochastic matrix) is denoted by \( A_i \). We call the set of fading states associated with the \( i^{th} \) block a macro-state and denote it by \( S_i \). The number of micro-states in \( S_i \) is denoted by \( N_i \). Throughout we assume that \( A_i \) has \( N_i \) distinct eigenvalues. We assume that the receiver knows the current micro-state of the channel whereas the sender only knows the current macro-state.

Let \( \pi^{(i)} \) be the stationary probability vector associated with \( A_i \), i.e.,

\[
\pi^{(i)} A_i = \pi^{(i)}.
\]

Define an \( M \times M \) matrix \( P \) as follows: the \((i,j)\) entry of \( P \) is given by

\[
P_{ij} = \sum_{k \in S_i, l \in S_j} \pi_k^{(i)} B_{kl}, \quad i \neq j,
\]

and

\[
P_{ii} = 1 - \sum_{j \neq i} P_{ij}.
\]

Note that \( P \) is also a stochastic matrix and let \( p \) be its stationary probability vector, i.e., \( p = pP \). We can interpret \( P \) as being the long-term transition probabilities among macro-states and \( p_i \) as approximating the long-term probability of being in \( S_i \), i.e., \( p_i(\epsilon) = p_i + O(\epsilon) \), where \( p_i(\epsilon) \) is the actual probability of being in micro-state \( i \).

Let \( T(n) \) be the random variable representing the micro-state at time \( n \) and \( S(n) \) be the random variable representing the macro-state at time \( n \). We generally use upper case letters to denote random variables and lower case letters to denote their sample values. Further, let \( G(T(n)) \) be the complex random variable corresponding to the signal attenuation at time \( n \). The received signal at time \( n \) is given by the random variable

\[
Y(n) = G(T(n))X(n) + W(n),
\]

where \( X(n) \) is the transmitted signal and \( W(n) \) is AWGN with variance \( \sigma^2 \).

We now state our main result.

Theorem 1 Define

\[
C := \max_{\{P(i)\}} \frac{1}{2} \sum_{i=1}^{M} \sum_{\epsilon \in S_i} \log \left( 1 + \frac{P(i)|G(\epsilon)|^2}{\sigma^2} \right) \pi_i^{(i)}.
\]

subject to \( \sum_{i=1}^{M} p_i(\epsilon)P(i) \leq P \), where \( P \) is the power constraint on the sender. Also, define

\[
C_{true}(\epsilon) = \lim_{n \to \infty} \frac{1}{n} \max_{P} I(X^n; Y^n, T^n),
\]

\[
p(x^n|x^k) = p(x^n|x^k), \text{ for all } k \geq n. \text{ Then, there exists } \epsilon^*(C) \text{ such that for } \epsilon \in (0, \epsilon^*(C))
\]

\[
C_{true}(\epsilon) = C + O(\epsilon \log(\epsilon) \log(-\epsilon \log(\epsilon))).
\]
Note that in this paper we speak of capacity and average mutual information interchangeably. The fact that the type of model we consider allows for capacity computations has recently been shown in [1].

2. Channel Capacity Using A Reduced-Order Model At the Transmitter

The difficulty in establishing the capacity of the channel lies in the fact that the macro-state transitions give us partial information about the current microstate. This information is difficult to quantify. However, as we will show, the longer we stay in a macrostate, this information becomes less relevant. In order to prove our result, we construct a channel whose capacity upper bounds the capacity of our channel. We provide a lower bound by choosing a specific input distribution, and then, we show that the difference between the upper bound and the lower bound is \( O(\epsilon \log(-\epsilon \log(\epsilon))) \).

Let us call our original channel \( C \) and let \( t_{\text{max}} \) be the channel state which corresponds to the highest SNR of all states in \( S \). The upper bound to the capacity of our channel \( C \) is obtained by constructing a new channel \( C_U \) as follows. Whenever there is a transition from one macro-state to another in \( C \), the channel \( C_U \) remains in the state \( t_{\text{max}} \) for time \( \tau \) until no macro-state transitions have occurred for \( \tau \) time in \( C \). After that time, \( C_U \) reverts to the macro-state that \( C \) is in. Moreover, the receiver is told what the micro-state was for channel \( C \) at the time of the last macro-state transition. Thus, under our definition, if a macro-state change occurs \( \frac{r}{2} \) time after the last macro-state change and the next macro-state change occurs more than \( \tau \) time after, then \( C_U \) will remain in state \( t_{\text{max}} \) for \( \frac{r}{2} \) time. At the end of that time, \( C_U \) reverts to behaving like \( C \), except that the the SCIS of \( C_U \) includes the micro-state at the time of the last macro-state change. Figure 1 illustrates a sample behavior for \( C \) and the corresponding behavior for \( C_U \). The capacity of \( C_U \) is higher than that of \( C \). During the intervals shown as shaded areas in Figure 1, the channel \( C_U \) has perfect RCSI and SCSI and, moreover, its SNR is the best possible SNR, which is given by the gain \( G(t_{\text{max}}) \) of the best micro-state \( t_{\text{max}} \) over all sets \( S_i \). Outside those shaded areas, \( C_U \) behaves as \( C \), except that it has better SCSI. The extra SCSI can only be advantageous in terms of mutual information, since the sender could always elect to disregard the extra information. Let \( t_k \) the beginning of the \( k^{th} \) white interval (i.e. interval where \( C_U \) behaves as \( C \)), we shall call an active interval, and \( \lambda_k \) be its duration.

A lower bound on the capacity is given by

\[
C = \max_{\{P(i)\}} \frac{1}{2} \sum_{i=1}^{M} p_i(\epsilon) \sum_{t \in S_i} \log \left( 1 + \frac{P(i)G(t)^2}{\sigma^2} \right) p_i^{(i)}(\epsilon),
\]

subject to \( \sum_{i=1}^{M} p_i(\epsilon) P(i) \leq P \), where \( p_i^{(i)}(\epsilon) \) and \( p_i(\epsilon) \) are the exact probabilities of being in micro-state \( t \) and macro-state \( S_i \), respectively. The lower bound holds by virtue of the fact that selecting a particular distribution for the input cannot yield a better result than maximization over all allowable input distributions. Specifically, we choose the input symbols to be independent zero-mean Gaussian random variables with variance \( P(i) \) in macro-state \( S \). The idea of using an upper bound which uses additional information and a lower bound which uses a specific input has been used in [3], although in a very different context.

2.1. Channel behavior in terms of \( \epsilon \).

In this section, we show that, if \( C \) stays in a macrostate \( S \) for some time \( L \) which is at least \( 2\pi + 2 \) time units, where \( \tau = O(\log(\epsilon)) \), then, for any time instant \( \tau < k < L - \tau \), the probability that \( C \) is in state \( i \in S \) conditioned on still being in \( S \) at time \( k \), is \( O(\epsilon) \) close to \( \pi^{(m)} \). Recall that \( \pi^{(m)} \) is the stationary probability associated with block \( A_i \). The motivation for showing this is as follows: the upper bound channel has additional sender side information at the time of each macro-state transition. However the results of this section will show that, after \( \tau \) units, the sender's estimate of the channel state is nearly independent of this additional side information. Owing to space limitations, we state the main result without a proof.

Theorem 2 Let \( t(0) = a \in S_m \) and define \( \beta_i^{m}(k) = \text{Prob}(t(k) = i|t_{k-1}^{L-1} = S_m, t_{k-k''}^{k''}) \) for all \( 0 < k', k, k'' > n \) and \( k'' > k \). Then, there exists \( K_m \) and \( \epsilon^* \) such that for all \( k, k'' > n \)

\[
\frac{-\log(\epsilon) + \log K_m}{-\log |\lambda_{2m(\epsilon)}|} < k < L - \frac{-\log(\epsilon) + \log K_m}{-\log |\lambda_{2m(\epsilon)}|},
\]

698
\[ |\pi_i^{(m)} - \beta_i(k)| = O(-\epsilon \log(\epsilon)), \text{ for all } i \in S_m. \]

### 2.2. Computation of capacities.

We first remark that

\[
I(X^n; Y^n, T^n | S^n) = I(X^n; T^n | S^n) + I(X^n; Y^n | T^n | S^n)
\]

\[
= h(Y^n | T^n) - h(X^n) \tag{4}
\]

Since the Xs are chosen on the basis of the current S only, we can use Hadamard's inequality and the fact that a Gaussian distribution maximizes entropy for a given covariance matrix to obtain the fact that, for our constraint on selecting the input distribution, the maximizing input distribution for the Xs is zero-mean Gaussian, with all the Xs mutually independent.

Let us consider C\(_U\). Its capacity can be upper-bounded by the sum of the capacities for two channels, C\(_{U1}\) and C\(_{U2}\), each of which has the same average input power constraint as our original channel C. Channel C\(_{U1}\) is created as follows: for every transition from one macro-state to another in C, the channel C\(_{U1}\) produces no output for time \(\tau\) until no macro-state transitions have occurred for \(\tau\) time in C. At all other times, which we call active intervals as for C\(_U\), C\(_{U1}\) follows the behavior of C. In addition to the SCSI of C, the SCSI of C\(_{U1}\) includes all the micro-states at all past macro-state transitions. Channel C\(_{U2}\) is constructed as follows: for every time the channel C\(_U\) experiences a macro-state transition, C\(_{U2}\) has an output for time \(\tau\). Moreover, while there is output, the channel behaves as a simple AWGN channel corresponding to the best possible t(n), which we denote t\(_{max}\). The idea of using an upper and lower bound to achieve capacity is similar to [2]; however, the upper bound channel that we use here is different.

**Lemma 1** C\(_{U1}\) = O(\(\epsilon \log(\epsilon)\)) \(\log(-\epsilon \log(\epsilon))\), where C\(_{U1}\) is the capacity of the channel C\(_U\).

Let us now consider C\(_U\) and obtain an upper bound to C\(_{U1}\). Let us define \(\Theta[k]\) to be the duration of a stay in a macro-state for the \((k - 1)\)th transition to a macro-state. Let \(\beta\) be the sum of the absolute values of the terms of B. From our model, since the micro-state transition matrix is of the form \(A + \epsilon B\), then for all k

\[
E[\Theta[k]] > \frac{1}{\beta \epsilon} \tag{5}
\]

Furthermore let us define \(\overline{\Theta}[n]\) to be the sample mean length of stay in a macro-state up until time n. Using the Weak Law of Large Numbers for decorrelating random variables applied to \(\Theta[k]\), we obtain that

\[
\lim_{n \to \infty} P\left(\overline{\Theta}[n] > \frac{1}{\beta \epsilon}\right) = 1. \tag{6}
\]

From (6), we obtain that, as \(n \to \infty\), with probability one, the time spent transmitting over C\(_U\) is at least O(\(\frac{1}{\epsilon}\)).

Let P represent a vector of power assignments over n symbols. Since we know that the maximizing distribution is Gaussian and independent from symbol to symbol, we can simply consider the problem of maximizing the mutual information over the set of power assignments. If we knew, for all the n symbols, what the SCSI is for the past, present and future, then we could only achieve a higher maximum. Conditioned on that information, we can select for a particular realization of the SCSI over n symbols a vector \(\text{P}_U\) which maximizes the average mutual information for the n symbols. The components of the vector \(\text{P}_U\) are \(\text{P}_U(l)\), which is the power used at time sample l. Let \(D_i(l)\) be the indicator function which takes the value 1 iff the macro-state at time i is \(S_l\) and 0 otherwise. The maximum average per symbol mutual information over n symbols subject to the average power constraint

\[
\sum_{i=1}^{n} \frac{\text{P}_U(l)}{n} \leq P, \text{ denoted by } MI(n), \text{ is given by}
\]

\[
\frac{1}{2n} \sum_{i=1}^{M} \sum_{l \in S} \sum_{i=1}^{n} D_i(l) \log \left(1 + \frac{\text{P}_U(l) | G(l) |^2}{\sigma^2}\right) \pi_i^{(l)}(l) \tag{7}
\]

where \(\pi_i^{(l)}(l)\) is the probability of being at time l in micro-state t of macro-state \(S_l\), which is the macro-state at time l, conditioned on our present and future SCSI over all n symbols. From our results in Theorem 2, we may write that

\[
MI(n) = \frac{1}{2n} \sum_{i=1}^{M} \sum_{l \in S} \sum_{i=1}^{n} D_i(l) \log \left(1 + \frac{\text{P}_U(l) | G(l) |^2}{\sigma^2}\right) \pi_i^{(l)} + O(-\epsilon \log(\epsilon)) \tag{8}
\]

subject to the average power constraint \(\sum_{i=1}^{n} \frac{\text{P}_U(l)}{n} \leq P\). For ease of notation, we may perform the following bijective mapping on (1, 2, ..., n):

\[
m : (1, 2, ..., n) \mapsto (1, 2, ..., n)
\]

\[
l \mapsto m(l) \text{ so that if } S(l) < S(j) \text{ then } m(l) < m(j) \text{ and if } S(l) = S(j), \text{ then } l < j \implies m(l) < m(j).
\]

Let \(m^{-1}\) be the inverse mapping of m. Let us define \(N_i = \sum_{l=1}^{n} D_i\). \(N_i\) is a random variable and we denote its sample values by \(n_i\). Using this definition, we may
rewrite (8) as

\[ MI(n) = \frac{1}{2n} \sum_{i=1}^{M} \sum_{l \in S_i} \sum_{t=1}^{n_i} \left[ D_i \left( M^{-1}(l) \right) \times \log \left( 1 + \frac{P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \pi_t^{(i)} \right] + O(-\epsilon \log(\epsilon)) \]

subject to the average power constraint

\[ \sum_{i=1}^{M} \sum_{t=n_{i-1}+1}^{n_i} P_U(m^{-1}(l)) \leq P, \]

where \( n_0 = 0 \). Let us denote \( N = \sum_{i=1}^{M} N_i \). We may write that

\[ C_U^1 \leq \lim_{n \to \infty} E[MI(n)] \]

where the expectation is over all realizations of macrostates over \( n \) symbols and where we are still subject to the average power constraint. We have from (9, 10) that

\[ C_U^1 \leq \lim_{n \to \infty} E_{N_1 \ldots N_M,S^r} \left[ \sum_{i=1}^{M} \frac{1}{2} \sum_{t \in S_i} \frac{N_i}{n N_i} \times D_i(m^{-1}(l)) \log \left( 1 + \frac{P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \pi_t^{(i)} \right] + O(-\epsilon \log(\epsilon)) \]

Using the concavity of \( \ln(1 + x) \) and Jensen's inequality,

\[ C_U^1 \leq \lim_{n \to \infty} E_{N_1 \ldots N_M,S^r} \left[ \sum_{i=1}^{M} \frac{1}{2} \sum_{t \in S_i} \frac{n_i}{n} D_i(m^{-1}(l)) \times \log \left( 1 + \frac{\sum_{l=n_{i-1}+1}^{n_i} P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \pi_t^{(i)} \right] + O(-\epsilon \log(\epsilon)) \]

subject to \( \sum_{i=1}^{M} \sum_{t=n_{i-1}+1}^{n_i} P_U(m^{-1}(l)) \leq P \) for every set of possible values of the \( N_i \)'s. We can rewrite the left-hand side of the constraint as

\[ \sum_{i=1}^{M} \sum_{l=n_{i-1}+1}^{n_i} P_U(m^{-1}(l)) \frac{n_i}{n} \]

The constraint can be weakened (thus giving a further upper bound) if we replace (13), which is a constraint for each realization, by its expected value over the \( S_i \)'s, conditioned in the values of the \( N_i \)'s. The left-hand side of the weakened constraint is:

\[ \sum_{i=1}^{M} E_{s^r} \left[ N_1 = \ldots = N_M = n \right] \left[ \sum_{l=n_{i-1}+1}^{n_i} D_i \left( m^{-1}(l) \right) P_U(m^{-1}(l)) \right] \frac{n_i}{n} \]

Moreover, in (12), we can pull the \( D_i \) into the log, since

\[ D_i \left( m^{-1}(l) \right) \log \left( 1 + \frac{\sum_{l=n_{i-1}+1}^{n_i} P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \]

\[ = \log \left( 1 + \frac{\sum_{l=n_{i-1}+1}^{n_i} D_i \left( m^{-1}(l) \right) P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right). \]

Using Jensen's inequality and relaxing our constraints further, we may finally obtain, after some manipulation, that

\[ C_U^1 \leq \lim_{n \to \infty} E_{N_1 \ldots N_M,S^r} \left[ \sum_{i=1}^{M} \frac{1}{2} \sum_{t \in S_i} \frac{n_i}{n} \times \log \left( 1 + \frac{P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \pi_t^{(i)} \right] + O(-\epsilon \log(\epsilon)) \]

\[ = \left[ \frac{1}{2} \sum_{i=1}^{M} \sum_{t \in S_i} p_i(\epsilon) \log \left( 1 + \frac{P_U(m^{-1}(l)))G(t)^2}{\sigma^2} \right) \pi_t^{(i)} \right] + O(-\epsilon \log(\epsilon)) \]

subject to \( \sum_{i=1}^{M} P_U(m^{-1}(l)) n_i \leq P \).

We may, yet again, weaken our constraint by replacing \( \frac{n_i}{n} \) by its expectation, \( p_i(\epsilon) \). This, together with (17), is the expression given in Theorem 1 by (2), to within \( O(-\epsilon \log(\epsilon)) \).

References

