Wideband Fading Channels with Feedback

Shashibhushan Borade Lizhong Zheng

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{spb,lizhong}@mit.edu

Abstract

We consider the flat fading Rayleigh channel at low SNR. With full CSI at the transmitter and the receiver, its capacity is shown to behave as $\log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$, as the transmit power denoted by SNR goes to zero. In fact, this capacity can be achieved with a single bit of CSI at the transmitter and without any receiver CSI. Then the capacity for the case of noisy CSI at the transmitter is found at low SNR. We then consider the case of block fading channel of large coherence interval T having unit delay feedback and no a priori CSI. If the only use of feedback is for power control, it is shown that the capacity behaves as $\log(T)\mathsf{SNR}$ in the limit of low SNR.

1 Introduction

It is well known that feedback does not improve the capacity of memoryless channels. Nonetheless, feedback can reduce the computational complexity significantly. For the case of channels with memory, feedback[1] can give significant capacity gains as well. In his paper [1], Massey had shown that the rate R achieved by a coding strategy in a channel is upper bounded by its directed information as follows:

$$R \leq \lim_{N\to\infty} \frac{I(\mathbf{x}^N \to \mathbf{y}^N)}{N}$$

$$\text{where} \quad I(\mathbf{x}^N \to \mathbf{y}^N) \triangleq \sum_{i=1}^{N} I(\mathbf{x}^i; \mathbf{y}_i \,|\mathbf{y}^{i-1})$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ denote the input $\mathbf{x}$ and output $\mathbf{y}$ of the channel at time $i \in \mathcal{Z}$, $\mathbf{x}^i$ denotes the vector of inputs $(\mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_i)$ and N is the codelength. Although the directed mutual information gives an analytical formula, optimizing it over the set of all input distributions and feedback schemes is a difficult optimization problem for most channels. Hence the feedback capacity and the optimal coding strategies for most channels with memory are not well understood. In this paper, we consider a simple case of feedback channels with memory, namely the flat fading channel. Moreover, we will assume a simple block

[1]The term *feedback* will mean perfect feedback with unit delay unless stated otherwise.

structure on the sequence of states. This structure has the same state within blocks of a fixed length (say T) and the states over different blocks are independent.

Compared to the feedback problem, utility of the channel state information (CSI) at the receiver and/or transmitter is understood better in many cases. We first consider the no feedback case of the Rayleigh flat fading channel

$$\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{y}$, $\mathbf{x}$, $\mathbf{h}$, and $\mathbf{w}$ are complex random variables. The additive white noise $\mathbf{w}$ is independent of all other random variables. We denote a complex circularly symmetric Gaussian random variable with mean m and variance σ^2 as $\mathcal{CN}(m, \sigma^2)$. In this notation, we have $\mathbf{h}, \mathbf{w} \sim \mathcal{CN}(0, 1)$. The transmit power is constrained as $\mathcal{E}\left[|\mathbf{x}|^2\right] \leq \mathsf{SNR}$.

Note that having full CSI at both ends converts this channel to a memoryless channel and its capacity is the same with or without feedback. Hence the capacity with full CSI is an upper bound to the capacity of the feedback channel[2] without any a priori CSI.

For the case of full CSI at the receiver and the transmitter, the capacity is given by water-filling over time [2]. At high SNR, the capacity improvement due to CSI at the transmitter is negligible. At low SNR on the other hand, the ratio of the capacity with and without transmitter CSI goes to infinity as SNR goes to zero. Thus CSI is of significant value at low SNR.

Although full CSI at both ends gives significant capacity gains, obtaining such CSI can be difficult at low SNR, potentially reducing the gains. We will therefore study some of these non-ideal cases.

First, even if the CSI is available to the receiver, the capacity of the feedback channel is sometimes limited. We are thus interested in the capacity gains with limited feedback [4]. We show that in the low SNR regime, a very limited feedback suffices to cause essentially the same capacity gain as the perfect feedback.

Second, obtaining receiver CSI is itself difficult in the low SNR regime. In the slow fading case, corresponds to a coherence time $T = \infty$, the channel can be estimated at the receiver with a vanishing fraction of the total energy spent in sending training signals, hence the CSI can be assumed to be available at the receiver at no cost. In the case with a finite coherence time, however, the cost of learning the channel is no longer negligible. We characterize in this paper how does the channel capacity approaches to that of the full CSI case as $T \to \infty$.

The rest of the paper is organized as follows. Next section discusses the capacity of this channel without feedback at low SNR in different cases of transmitter/receiver CSI. The third section studies the feedback channel. It studies the case of large coherence interval T at low SNR. Some discussion follows in the end.

2 Value of the Channel State Information

For the full CSI case, the energy efficiency i.e. the capacity per unit energy is known to go to infinity as the SNR goes to zero [3]. We now state the following lemma which shows how fast the energy efficiency goes to infinity as SNR goes to zero.

[2]The block fading Rayleigh channel with feedback, coherence interval T, and no a priori CSI is simply called as the feedback channel henceforth.

Lemma 1 *With full CSI at the transmitter and receiver, capacity* $C_{TR}(\mathsf{SNR})$ *of the Rayleigh fading channel in (1) satisfies*

$$\lim_{\mathsf{SNR}\to 0} \frac{C_{TR}(\mathsf{SNR})}{\log(\frac{1}{\mathsf{SNR}})\cdot \mathsf{SNR}} = 1$$

We will use the notation $f(\mathsf{SNR}) \approx g(\mathsf{SNR})$ as a shorthand[3] for $\lim_{\mathsf{SNR}\to 0} \frac{f(\mathsf{SNR})}{g(\mathsf{SNR})} = 1$. For example, the above lemma can be written as $C_{TR}(\mathsf{SNR}) \approx \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$.

Proof: For a channel strength of $t \triangleq |\mathbf{h}|^2$, the water-filling solution puts $\left(\frac{1}{\alpha} - \frac{1}{t}\right)^+$ amount of power[4], where α is the channel threshold chosen to satisfy the average power constraint

$$\mathsf{SNR} = \int_0^\infty \left(\frac{1}{\alpha} - \frac{1}{t}\right)^+ e^{-t}dt = \int_\alpha^\infty \left(\frac{1}{\alpha} - \frac{1}{t}\right) e^{-t}dt \tag{2}$$

because t has an exponential distribution of mean 1. The channel capacity $C_{TR}(\mathsf{SNR})$ in terms of the channel threshold α is given by

$$C_{Th}(\alpha) \triangleq \int_\alpha^\infty \log\left(\frac{t}{\alpha}\right) e^{-t}\, dt \tag{3}$$

We first show that the channel threshold α satisfying Eq. (2) can be 'tightly' approximated by $\log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}})$ as SNR goes to 0. The RHS of Eq. (2) can be written as $P_{Th}(\alpha)$, where $P_{Th}(a)$ is defined as

$$\begin{aligned} P_{Th}(a) &\triangleq \int_a^\infty \left(\frac{1}{a} - \frac{1}{t}\right) e^{-t}dt \qquad (4)\\ &= \int_a^\infty \frac{e^{-t}}{t^2}dt \qquad \text{(by integration by parts)} \qquad (5)\end{aligned}$$

Note that for all $a > 0$,

$$P_L(a) \triangleq \frac{e^{-a}}{a^2} - \frac{2e^{-a}}{a^3} \le P_{Th}(a) \le \frac{e^{-a}}{a^2} \triangleq P_U(a)$$

Now substitute $a_U = \log\left(\frac{1+\delta}{\mathsf{SNR}}\right) - 2\log\log(\frac{1}{\mathsf{SNR}})$ in $P_U(a)$ for some $\delta > 0$. It satisfies

$$\begin{aligned} \lim_{\mathsf{SNR}\to 0} \frac{P_U(a_U)}{\mathsf{SNR}} &= \frac{1}{1+\delta}\\ \text{Hence} \quad \lim_{\mathsf{SNR}\to 0} \frac{P_{Th}(a_U)}{\mathsf{SNR}} &\le \frac{1}{1+\delta}\end{aligned}$$

Now substitute $a_L \triangleq \log\left(\frac{1}{(1+\delta)\mathsf{SNR}}\right) - 2\log\log(\frac{1}{\mathsf{SNR}})$ in $P_L(a)$. It satisfies,

$$\begin{aligned} \lim_{\mathsf{SNR}\to 0} \frac{P_L(a_L)}{\mathsf{SNR}} &= 1+\delta\\ \text{Hence} \quad \lim_{\mathsf{SNR}\to 0} \frac{P_{Th}(a_L)}{\mathsf{SNR}} &\ge 1+\delta\end{aligned}$$

[3] $f(\mathsf{SNR}) \,\tilde{>}\, g(\mathsf{SNR})$ and $f(\mathsf{SNR}) \,\tilde{<}\, g(\mathsf{SNR})$ are similarly defined.

[4] The function $(x)^+ = x$ for positive x and zero otherwise.

Considering that $P_{Th}(a)$ is a strictly decreasing function, α should lie between a_L and a_U. Both a_L and a_U are equal to $\log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}})$ up to an additive constant. Hence α also equals $\log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}})$ up to an additive constant as $\mathsf{SNR} \to 0$.

Noting that $C_{Th}(a)$ is a decreasing function and $\frac{e^{-a}}{a} - \frac{e^{-a}}{a^2} \leq C_{Th}(a) \leq \frac{e^{-a}}{a}$ gives

$$\frac{e^{-a_U}}{a_U} - \frac{e^{-a_U}}{a_U^2} \leq \; C_{Th}(a_U) \leq \; C_{Th}(\alpha) \leq \; C_{Th}(a_L) \leq \; \frac{e^{-a_L}}{a_L}$$

We can verify that

$$\lim_{\mathsf{SNR}\to 0} \frac{e^{-a_L}/a_L}{\log(1/\mathsf{SNR})\mathsf{SNR}} = 1+\delta \quad \text{and} \quad \lim_{\mathsf{SNR}\to 0} \frac{e^{-a_U}/a_U - 2e^{-a_U}/a_U^2}{\log(1/\mathsf{SNR})\mathsf{SNR}} = \frac{1}{1+\delta}$$

$$\text{Hence} \quad \frac{1}{1+\delta} \leq \lim_{\mathsf{SNR}\to 0} \frac{C_{Th}(\alpha)}{\log(1/\mathsf{SNR})\mathsf{SNR}} \leq 1+\delta$$

The lemma follows as arbitrarily small δ can be chosen.

Now we show that the same result follows when a simple On-Off power control is used instead of the optimal water-filling. This power strategy either transmits a single non-zero power level or no power depending on the channel strength.

Lemma 2 *In a Rayleigh fading channel with full CSI at both ends, the achievable rate* $R(\mathsf{SNR})$ *by an On-Off power allocation (instead of the optimal water-filling) satisfies*

$$\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\log(\frac{1}{\mathsf{SNR}}) \cdot \mathsf{SNR}} = 1$$

Proof: An On-Off power control strategy is employed which transmits at a non-zero power level SNR' for all $t \triangleq |\mathbf{h}|^2 \geq \log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}}) \triangleq \Theta$ and keeps silent otherwise. Since t is exponentially distributed with mean 1, the non-zero power level $\mathsf{SNR}' = \mathsf{SNR}/e^{-\Theta}$. The achievable rate with this strategy equals

$$\begin{aligned} R(\mathsf{SNR}) &= \int_{\Theta}^{\infty} \log\left(1 + \frac{t\mathsf{SNR}}{e^{-\Theta}}\right) e^{-t} dt \\ &\geq \int_{\Theta}^{\infty} \log\left(1 + \frac{\Theta\mathsf{SNR}}{e^{-\Theta}}\right) e^{-t} dt = e^{-\Theta}\log(1 + e^{\Theta}\Theta\ \mathsf{SNR}) \end{aligned}$$

Note that $e^{\Theta}\Theta\ \mathsf{SNR}$ goes to zero with SNR. Hence, $R(\mathsf{SNR}) \approx e^{-\Theta}(e^{\Theta}\Theta\ \mathsf{SNR}) = \Theta\mathsf{SNR}$. Thus the channel threshold Θ directly gives the energy efficiency at low SNR. Its easy to see that the same proof also holds for $\Theta = \log(\frac{1}{\mathsf{SNR}}) - (1+\epsilon)\log\log(\frac{1}{\mathsf{SNR}})$ for any $\epsilon > 0$.

Note that whenever non-zero power is transmitted in the above scheme, the received SNR (i.e. $|\mathbf{h}|^2 e^{\Theta}\mathsf{SNR}$) almost always goes to zero with SNR. This explains the capacity achieving nature of this strategy. The channel is almost always in the low SNR regime, where the rate and received energy have a linear relationship. This gives an easy guiding principle: stay in the linear region as much as possible.

The above power allocation only needs one bit of CSI at the transmitter. This bit indicated whether or not the channel is better than the threshold Θ. Thus at low SNR, the capacity with only one bit of CSI at the transmitter is the same as that with full CSI at the transmitter. This is the limited feedback channel, where the receiver has full CSI and the transmitter has a quantized CSI of finite precision. Thus at low SNR, capacity of the limited feedback channel equals the capacity of the channel full transmitter CSI.

Remark 1: Lemma 1 and 2 hold true even when no CSI is available at the receiver. The proof is omitted here but it follows on similar lines of Theorem 1 in [5], where the receiver employs energy detection for decoding.

Remark 2: For the channels in Lemma 1 and 2, [3] previously showed that the ratio of the capacity and SNR goes to infinity with or without any receiver CSI.

2.1 Partial Transmitter CSI

Up to this point, we studied the channel capacity when the transmitter has noiseless CSI. Even the one bit CSI case assumed no noise, that is, the channel was indeed better or worse than the threshold if the CSI conveyed so. However it may be impossible to obtain such noiseless CSI. For example, if the channel is trained with energy E for obtaining CSI, the channel estimate is noisy for any finite training energy $E > 0$. The actual channel realization $\mathbf{h}$ in this case is given by the sum of two components: the channel estimate $\mathbf{g}$ and the estimation error $\mathbf{f}$. The channel equation in this case is given by

$$\mathbf{y} = (\mathbf{g} + \mathbf{f})\mathbf{x} + \mathbf{w} \tag{6}$$

For the Rayleigh, the MMSE channel estimate and the error are independent of each other. We assume the channel estimate $\mathbf{g} \sim \mathcal{CN}(0, \beta)$ and the error $\mathbf{f}$ is an independent $\mathcal{CN}(0, 1-\beta)$. Note that this can correspond to training with energy E, such that the MMSE error variance $1/(1+E)$ equals $1-\beta$.

We know that $\beta = 1$ corresponds to the full CSI case and $\beta = 0$ is the no CSI case. At low SNR, their channel capacities are essentially $\log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$ and SNR, respectively. We expect intuitively that the capacity should increase with β in some manner. The following theorem tells how exactly it increases with β at low SNR. It bridges the two extremes cases of $\beta = 0$ and $\beta = 1$.

Theorem 3 *Consider the channel* $\mathbf{y} = (\mathbf{g} + \mathbf{f})\mathbf{x} + \mathbf{w}$*, where* $\mathbf{g}$ *and* $\mathbf{f}$ *are independent Rayleigh random variable with variance* β *and* $\beta' = 1 - \beta$*, respectively. The transmitter only knows* $\mathbf{g}$ *and the receiver knows both* $\mathbf{g}$ *and* $\mathbf{f}$*. The capacity* $C_\beta(\mathsf{SNR})$ *of this channel for any fixed* $\beta \in (0, 1]$ *satisfies*

$$\lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\beta \cdot \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}} = 1 \quad i.e. \quad C_\beta(\mathsf{SNR}) \approx \beta \cdot \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$$

Thus the capacity is essentially reduced by a factor of β *due to the noisy channel estimate.*

Proof: We first prove an upper bound on the capacity of this channel.

$$C_\beta(\mathsf{SNR}) = \mathcal{E}_{\mathbf{g},\mathbf{f}} \left[\log(1 + |\mathbf{g} + \mathbf{f}|^2 P(\mathbf{g}))\right]$$

because the transmitter only knows $\mathbf{g}$, which is independent of $\mathbf{f}$, the transmit power is a function of $\mathbf{g}$ only (denoted by $P(\mathbf{g})$). Jensen's inequality and independence of $\mathbf{g}$ and $\mathbf{f}$ implies

$$\begin{aligned} C_\beta(\mathsf{SNR}) &\leq \mathcal{E}_{\mathbf{g}} \left[\log\left(1 + \mathcal{E}_{\mathbf{f}}\left[|\mathbf{g} + \mathbf{f}|^2 P(\mathbf{g})\right]\right)\right] \\ &= \mathcal{E}_{\mathbf{g}} \left[\log\left(1 + (|\mathbf{g}|^2 + \beta')P(\mathbf{g})\right)\right] \\ &\leq \mathcal{E}_{\mathbf{g}} \left[\log\left(1 + \beta' P(\mathbf{g})\right)\right] + \mathcal{E}_{\mathbf{g}} \left[\log\left(1 + |\mathbf{g}|^2 P(\mathbf{g})\right)\right] \\ &\leq \log(1 + \beta' \mathcal{E}_{\mathbf{g}}\left[P(\mathbf{g})\right]) + \mathcal{E}_{\mathbf{g}} \left[\log\left(1 + |\mathbf{g}|^2 P(\mathbf{g})\right)\right] \end{aligned}$$

The first term above is the capacity of AWGN channel with signal to noise ratio of $\beta' \mathcal{E}_{\mathbf{g}}[P(\mathbf{g})] = \beta' \mathsf{SNR}$. The second term is the capacity of the unit variance Rayleigh fading channel with full CSI at both ends and transmit SNR of $\beta\mathsf{SNR}$. Now applying Lemma 1 implies

$$\lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\beta'\mathsf{SNR} + \beta \cdot \log\left(\frac{1}{\beta\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1 \tag{7}$$

For any fixed $\beta \in (0, 1]$, this implies

$$\lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\mathsf{SNR} + \beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1 \tag{8}$$

$$\text{and} \qquad \lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1 \tag{9}$$

For the lower bound to the capacity, we give an achievable scheme as follows. This is very similar to the achievable scheme for the full CSI case (i.e. Lemma 2). We transmit uniform power when $|\mathbf{g}|^2$ is greater than a threshold which is β times the threshold for the full CSI case. That is, we transmit only when $t_g \triangleq |\mathbf{g}|^2 \geq \beta\Theta \triangleq \Theta_\beta$. Since t_g is an exponential random variable with mean β, the probability of $t_g \geq \Theta_\beta$ is given by $\exp(-\Theta_\beta/\beta)$. Hence the transmit power when $t_g \geq \Theta_\beta$ is given by $e^{\Theta_\beta/\beta}\mathsf{SNR}$. The achievable rate of this power allocation is

$$\begin{aligned}
R(\mathsf{SNR}) &\geq \mathcal{E}_{|\mathbf{g}|^2 \geq \Theta_\beta,\ \mathbf{f}}\left[\log\left(1 + |\mathbf{g}+\mathbf{f}|^2\, e^{\Theta_\beta/\beta}\,\mathsf{SNR}\right)\right] \\
&\geq \mathcal{E}_{|\mathbf{g}|^2 \geq \Theta_\beta,\ \mathbf{f}}\left[\log\left(1 + (|\mathbf{g}| - |\mathbf{f}|)^2\, e^{\Theta_\beta/\beta}\,\mathsf{SNR}\right)\right] \\
&\geq \mathcal{E}_{|\mathbf{g}|^2 \geq \Theta_\beta,\ |\mathbf{f}|<A}\left[\log\left(1 + (|\mathbf{g}| - A|)^2\, e^{\Theta_\beta/\beta}\,\mathsf{SNR}\right)\right] \quad \text{(for some fixed } A > 0\text{)} \\
&= (1 - e^{A/\beta'}) \cdot e^{-\Theta_\beta/\beta} \cdot \log\left(1 + |\sqrt{\Theta_\beta} - A|^2\, e^{\Theta_\beta/\beta}\,\mathsf{SNR}\right)
\end{aligned}$$

It is easy to check that $|\sqrt{\Theta_\beta} - A|^2 e^{\Theta_\beta/\beta}\mathsf{SNR}$ goes to zero with SNR for our choice of Θ_β. Moreover noting that A is fixed and negligeble compared to Θ_β which goes to infinity,

$$\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} = \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\Theta_\beta\mathsf{SNR}} = (1 - e^{A/\beta'}) \cdot 1$$

We can bring this lower bound arbitrarily close to 1 by choosing large enough A. Thus for any $\epsilon > 0$ we get

$$1 - \epsilon \leq \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq \lim_{\mathsf{SNR}\to 0} \frac{C_\beta}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1$$

The theorem is proved because arbitrarily small ϵ can be chosen.

On similar lines of Lemma 2, this achievability proof also shows that only 1 bit about $\mathbf{g}$ is needed at the transmitter. This bit should indicate whether or not $|g|^2$ is greater than the threshold Θ_β.

Remark 3: Note the probability $\exp(-\Theta_\beta/\beta)$ of transmitting non-zero power is the same as that in the full CSI case since $\Theta_\beta = \beta\Theta$. Thus although the fraction of channels used is the same as the full CSI case, the noise in the transmitter CSI reduces the capacity by a factor of β. Note again that the channel threshold Θ_β directly gives the energy efficiency.

Remark 4: As in the case of Remark 1, above theorem holds true even when the receiver has no CSI at all. The proof again follows on similar lines of Theorem 1 in [5], which uses energy detection at the receiver. Thus in all cases of transmitter CSI we have discussed so far, the ratio of capacities with or without any receiver CSI goes to 1 as SNR goes to 0. This property can be shown to hold for all fading distributions with an exponential tail.

Remark 5: The upper bound in Eq. (8) also holds when β goes to 0 with SNR instead of being fixed. We omit the proof which follows on the similar lines of the fixed β case. Hence for any β (fixed or decreasing with SNR), Eq. (8) implies that the capacity of the channel $\mathbf{y} = (\mathbf{g}+\mathbf{f})\mathbf{x}+\mathbf{w}$ at low SNR is worse than the sum capacity of the following two parallel channels. One channel is the unit variance Rayleigh channel with transmit power of SNR and no transmitter CSI. The other is a channel of capacity $\beta \log\left(\frac{1}{\mathsf{SNR}}\right) \mathsf{SNR}$ at low SNR.

3 Feedback Case for Block Fading

Now consider the block fading case with coherence interval T. Intuitively, we expect that the channel capacity should increase with the coherence time. This section studies how exactly it increases with T in the limit of large T. For this purpose of analyzing large T case, [6] considered the case when the coherence time T can be written as some negative power of SNR. On similar lines but more generally, we consider the case when the coherence interval T can be written as some function of SNR, which goes to infinity when SNR goes to zero. For example, T can be $\mathsf{SNR}^{-0.5}$ or $\log(1/\mathsf{SNR})$.

Theorem 4 *For the block fading Rayleigh channel with feedback, a rate $R(\mathsf{SNR})$ satisfying*

$$\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\log T \cdot \mathsf{SNR}} = 1$$

is achievable for any T going to infinity with SNR going to 0 such that $\log T \overset{\sim}{<} \log(\frac{1}{\mathsf{SNR}})$.

Proof: We train one out of every $\mathsf{SNR}^{-1}T^{-1+\epsilon}$ channel blocks for a fixed $\epsilon > 0$. The positions of the training blocks are predetermined and conveyed to both the ends. No energy is transmitted for the remaining blocks. The training signals are of fixed energy $E > 0$ and the training spans only the first symbol of that block. The remaining $T-1$ blocks are used for communication only. For each trained block, the total energy accumulated over $\mathsf{SNR}^{-1}T^{-1+\epsilon}$ blocks equals $E_{\text{Total}} = \mathsf{SNR}^{-1}T^{-1+\epsilon} \cdot T\mathsf{SNR} = T^{\epsilon}$. The fraction of this total energy used for training equals E/T^{ϵ} and can be ignored at low SNR because E is fixed and T goes to infinity with vanishing SNR. Hence the average SNR available in the trained blocks equals $\mathsf{SNR}' = E_{\text{Total}}/T = T^{\epsilon-1}$. Recall that a training energy of E corresponds to $\beta = \frac{E}{1+E}$ in Theorem 3. Applying Theorem 3 with $\beta = \frac{E}{1+E}$ implies that essentially a rate of $\beta\mathsf{SNR}'\log(1/\mathsf{SNR}')$ is achieved in every trained block. However since only $\mathsf{SNR}\ T^{1-\epsilon}$ fraction of the blocks are trained and only $T-1$ symbols are remaining for communication after training, the overall rate achieved satisfies

$$\begin{aligned}
\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\mathsf{SNR}\ T^{1-\epsilon} \cdot \beta\mathsf{SNR}'\log(1/\mathsf{SNR}')} &= 1 \lim_{\mathsf{SNR}\to 0} \frac{T-1}{T} = 1 \quad (\text{as } T \to \infty \text{ with } \mathsf{SNR} \to 0) \\
\Rightarrow \quad \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\mathsf{SNR}\log(T)} &= (1-\epsilon)\cdot\beta = (1-\epsilon)\cdot\frac{E}{1+E}
\end{aligned}$$

Choosing arbitrarily small ϵ and arbitrarily large training energy E yields the proof.

To interpret this result, assume that unit energy is needed to train a channel block perfectly. Also assume that 1 out of every $1/\delta$ trained blocks is 'good' i.e. above the channel threshold. This channel threshold should be $\log(1/\delta)$. Thus $1/\delta$ training energy should be spent for finding one good channel. Assume that the received SNR in the good block is given by ρ, which is much smaller than 1 to ensure a linear rate in received power. Hence the total transmit power in the good block is essentially given by $\frac{\rho}{\log(1/\delta)}T$. For the training to be effectively free, the training energy should be a small fraction of the communication energy. Since all the communication is done in the 'good' channel (see Fig. 1), the energy $\frac{\rho T}{\log(1/\delta)}$ transmitted in the good channel should be much larger than $1/\delta$. Thus δ should be larger than $1/T$. Hence the channel threshold $\log(1/\delta)$ i.e. the energy efficiency cannot be larger than $\log T$.

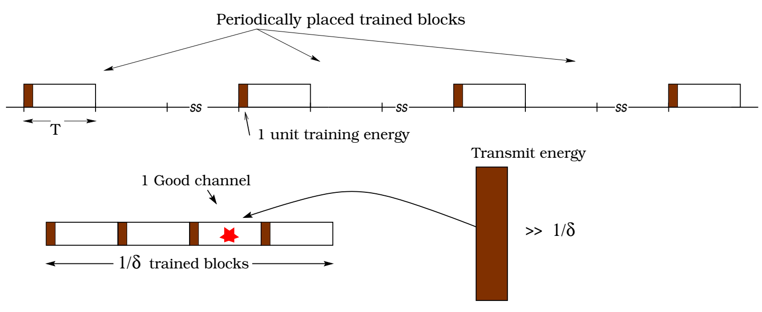

Figure 1: Training scheme for achievability

Note that very few bits of feedback are needed on average in this scheme. Feedback is only needed for the trained channels blocks. Moreover, as seen in the previous section, only one bit of feedback about the trained channel is good enough for achieving its capacity at low SNR. This bit should indicate whether or not the channel estimate is better than the threshold Θ_β. Thus on average a feedback of $\mathsf{SNR}\ T^{1-\epsilon}$ bits per block or $\mathsf{SNR}\ T^{1-\epsilon}/T = \mathsf{SNR}\ T^{-\epsilon}$ bits per symbol is required. Thus a very weak reverse channel is sufficient for this scheme.

Remark 6: The same scheme of training rarely with arbitrarily good quality can be applied to a channel with finite support fading whose maximum channel strength equals $|h|^2_{\max}$. If coherence interval T goes to infinity with SNR going to zero, this scheme achieves the capacity of this channel with full CSI at both ends. Equivalently, its capacity satisfies $C(\mathsf{SNR}) \approx |h|^2_{\max}\mathsf{SNR}$.

Remark 7: Note that when $\log T \approx \log(\frac{1}{\mathsf{SNR}})$, the achievable rate in Theorem 4 equals the capacity with full transmitter CSI. Hence the above strategy achieves the capacity if $\log(T) \overset{\sim}{>} \log(\frac{1}{\mathsf{SNR}})$. For smaller coherence intervals i.e. when $\log T \overset{\sim}{<} \log(\frac{1}{\mathsf{SNR}})$, we need a tighter upper bound than the capacity with full CSI. Assuming the transmitter only uses the feedback for adjusting the transmitted power, the following theorem gives an upper bound to the achievable rate. It matches with the achievable rate in Theorem 4 when $\log T \overset{\sim}{<} \log(\frac{1}{\mathsf{SNR}})$.

Theorem 5 *If the only use of feedback is for adjusting the transmit power, the capacity of the block fading Rayleigh channel with feedback satisfies*

$$\lim_{\mathsf{SNR}\to 0} \frac{C(\mathsf{SNR})}{\log T \cdot \mathsf{SNR}} \leq 1$$

for any T going to infinity with SNR *going to* 0.

Proof: Assume a genie which provides us with an extra parallel channel called the training channel for channel state estimation. This channel always takes the same value as the original channel. It also provides is with additional transmit power of SNR for this channel. Since the quality of the channel estimation only depends on the training energy, we may assume that all the training is done in the first symbol of the fading block of the training channel. Moreover, we assume that the estimate from the training channel is available to the communication channel just before its corresponding block starts.

Another genie tells the exact channel $\mathbf{h}$ to the transmitter when the training energy $E \geq 1$. Now we prove that the average power available for training should be only used for training perfectly i.e. training with energy 1.

By Remark 5, the capacity of a Rayleigh channel with $(1-\beta)$ estimation error at the transmitter is upper bounded by the $\mathsf{SNR} + \beta \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$ at low SNR. Now consider a distribution of training energies on the training channel. Say training with energy E_j is performed with probability p_j and the communication SNR allocated for this training level equals SNR_j. Assume that non-perfect training is also possible in this distribution, that is, some $0 < E_i < 1$ has $p_i > 0$. Since $\beta_i = \frac{E_i}{1+E_i}$, this training energy yields a rate of

$$p_i R_i(\mathsf{SNR}_i) \approx p_i \cdot \left(\mathsf{SNR}_i + \frac{E_i}{1+E_i}\log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i\right) \leq p_i \cdot \left(\mathsf{SNR}_i + E_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i\right) \quad (10)$$

The above inequality implies the training with $E_i < 1$ with probability p_i is worse than training perfectly with $E = 1$ with probability $p_i E_i$. The total training energy in both the schemes is the same. Thus any distribution of training energies can be improved (in terms of rate) by shifting all the imperfect trainings to perfect trainings with a lower probability. If p_0 indicates the probability of 0 training energy, the total achievable rate of the original suboptimal training distribution is upper bounded from Eq. (10) as

$$\begin{aligned} \sum_i p_i R_i(\mathsf{SNR}_i) &\stackrel{\sim}{<} p_0 \mathsf{SNR}_0 + \sum_{i\neq 0} p_i \cdot \left(\mathsf{SNR}_i + E_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i\right) \\ &\leq \sum_i p_i \mathsf{SNR}_i + \sum_{i\neq 0} p_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i \quad \text{as all } E_i \leq 1 \end{aligned}$$

Note that the average transmit power given by $\sum_i p_i \mathsf{SNR}_i$ at most equals SNR. Noting that $x\log(1/x)$ is a concave increasing function of x gives an upper bound on the RHS above as

$$\text{RHS} \leq \mathsf{SNR} + \mathsf{SNR}\log(\frac{1-p_0}{\mathsf{SNR}}) \ = \mathsf{SNR} + \mathsf{SNR}\log(\frac{p_1}{\mathsf{SNR}}) \ \text{ where } p_1 = 1 - p_0 \quad (11)$$

This corresponds to the fact that time-sharing between various codes at power SNR_i is not better than using one single code with the combined power.

We have proved that only perfect training should be performed which takes unit energy. The maximum probability p_1 of this perfect training is given by the total training energy constraint $(p_1 \cdot 1) \leq T\mathsf{SNR}$. Since the upper bound in Eq. 11 is increasing in p_1, we choose the maximum possible p_1. Hence the capacity of this channel satisfies

$$C(\mathsf{SNR}) \stackrel{\sim}{<} \mathsf{SNR} + \mathsf{SNR}\log(\frac{T\mathsf{SNR}}{\mathsf{SNR}}) = (1 + \log T)\mathsf{SNR}$$

Note that the total power used for this proof is essentially three times available power. The training channel uses SNR power, communication over trained blocks uses SNR and that over untrained blocks also uses roughly SNR power.

4 Discussion

Some engineering insights we obtained for the block fading feedback channel at low SNR are summarized: 1) The received SNR should be small as often as possible to take advantage of the linear rate-power behavior. 2) Whenever performed, the training should be almost perfect. 3) The energy spent in training should be a small fraction of the total available energy. 4) All the communication should be done during the trained blocks.

Finally, we remark that a one step Gauss-Markov fading with innovation rate equal to $\xi = \mathsf{SNR}^d$ can be converted to a block fading of essentially a coherence interval of SNR^{-d}. This ignores the correlation between adjacent blocks. By Theorem 4, a rate of $\log(1/\xi)\mathsf{SNR}$ can be achieved at low SNR. However, optimality of such conversion to block fading is not known.

Acknowledgements

The authors wish to acknowledge the helpful discussions with Robert Gallager, Sibi Raj and Emre Telatar.

References

[1] J. Massey, "Causality, feedback and directed information," *Proc. IEEE ISITA*, pp. 303–305, 1990.

[2] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. on Information Theory*, Vol. 43, No. 6, pp. 1986–1992, Nov. 1997.

[3] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. on Information Theory*, vol. 48, pp. 1319–1342, June 2002.

[4] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the Value of Limited Feedback for MIMO Channels?," *IEEE Communications Magazine*, Oct. 2003.

[5] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. on Information Theory*, vol. 46, pp. 1384–1400, July 2000.

[6] L. Zheng, D. N. C. Tse and M. Medard, "Channel coherence in the low SNR regime," *Proc. Allerton Conference*, October 2004.