Wideband Fading Channels With Feedback

Shashi Borade, *Member, IEEE*, and Lizhong Zheng, *Member, IEEE*

Abstract—The Rayleigh flat fading channel at low SNR is considered. With full channel state information (CSI) at the transmitter and receiver, its capacity is shown to be essentially $\mathsf{SNR}\log(1/\mathsf{SNR})$ nats/symbol, as SNR goes to zero. In fact, this rate can be achieved with a just one bit of CSI at the transmitter (per fading realization) and with no receiver CSI. The capacity for the case of noisy transmitter CSI is also found. Then a Rayleigh block fading channel of coherence interval $T \leq 1/\mathsf{SNR}$ is considered[1] which has causal feedback and no *a priori* CSI. A training based scheme is proposed for such channels, which achieves a rate of $\mathsf{SNR}\log T$ nats/symbol in the limit of small SNR and large T. Thus, when coherence interval T is of the order $1/\mathsf{SNR}$, without any *a priori* CSI at either end, the capacity with full CSI at both ends is achievable. For smaller values of T, a rate of $\mathsf{SNR}\log T$ nats/symbol is shown to be achievable.

Index Terms—Block fading, channel state information, coherence interval, fading channel, feedback capacity, low SNR, noisy CSI, wideband.

I. INTRODUCTION

FEEDBACK does not improve the capacity of memoryless channels. However, for channels with memory, feedback[2] can enhance the capacity. Based on the results in Massey's beautiful paper [1] and Fano's inequality, the rate[3] R of a code of length N can be upper bounded by its directed information $I(\mathbf{x}_1^N \to \mathbf{y}_1^N)$ and its error probability P_e^N as follows:

$$R \leq \sup_{\mathcal{E}_N} \frac{I(\mathbf{x}_1^N \to \mathbf{y}_1^N) + 1 + P_e^N N R}{N} \quad \text{where}$$

$$I(\mathbf{x}_1^N \to \mathbf{y}_1^N) \equiv \sum_{i=1}^{N} I(\mathbf{x}_1^i; \mathbf{y}_i \,|\, \mathbf{y}_1^{i-1}) \tag{1}$$

[1]A block fading channel of coherence interval T means that the channel state remains constant for each block T consecutive symbols. At the end of these T symbols, the channel state takes a new realization independently of its history. Each such channel state realization is Rayleigh distributed.

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

[3]Unless stated otherwise, the units of rate are nats/symbol and $\log(\cdot)$ denotes the natural logarithm.

where $\mathcal{E}_N$ denotes the set of all feedback codes of length N, $\mathbf{x}_i$ and $\mathbf{y}_i$ denote the input $\mathbf{x}$ and output $\mathbf{y}$ of the channel at time $i \in \mathcal{Z}$, and $\mathbf{x}_i^j$ denotes the sequence of inputs $(\mathbf{x}_i, \mathbf{x}_{i+1} \cdots \mathbf{x}_j)$. For reliable communication, the error probability P_e^N vanishes to 0 as N tends to infinity and the above upper bound becomes

$$R \leq \limsup_{N \to \infty} \sup_{\mathcal{E}_N} \frac{I(\mathbf{x}_1^N \to \mathbf{y}_1^N)}{N}. \tag{2}$$

Although the directed information gives an analytical expression, its optimization over all feedback codes is a difficult problem for most channels. Moreover, this optimization needs to be done for all N for analyzing feedback capacity. Hence the feedback capacity and optimal coding strategies for most channels with memory are unknown, besides few notable exceptions such as [2]–[4].

This paper considers a simple feedback channel with memory, namely the Rayleigh flat fading channel. Moreover, we will assume a simple block 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, channels with memory are better understood if the channel state information (CSI) is already available at the receiver and/or transmitter[4]. We first consider the no feedback case of the channel of interest, the Rayleigh flat fading channel

$$\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{w} \tag{3}$$

where $\mathbf{y}$, $\mathbf{x}$, $\mathbf{h}$, and $\mathbf{w}$ are complex random variables. The channel state $\mathbf{h}$ is Rayleigh distributed and the additive white noise $\mathbf{w}$ is independent of all other random variables. We denote a zero mean complex circularly symmetric Gaussian random variable with variance σ^2 as $\mathcal{CN}(0, \sigma^2)$, so we have $\mathbf{h}, \mathbf{w} \sim \mathcal{CN}(0,1)$. The transmit power is constrained as $\mathsf{E}\left[|\mathbf{x}|^2\right] \leq \mathsf{SNR}$, where $\mathsf{E}[\cdot]$ denotes expectation. In Section II, the channel state $\mathbf{h}$ is assumed to change i.i.d. for each symbol. In Section III, we assume a i.i.d. block fading model for $\mathbf{h}$, where remains constant for T symbols and then changes in an i.i.d. manner.

For the case of full CSI (knowledge of $\mathbf{h}$) at the receiver and transmitter, the capacity is given by water-filling over time [5]. At high SNR, capacity improvement due to such water-filling is negligible. At low SNR on the other hand, the ratio of the capacity with and without water-filling goes to infinity in the limit of small SNR [6]. Thus full transmitter CSI can provide significant capacity gains at low SNR through. It is not clear whether full CSI is necessary for achieving this significant gain

[4]We denote channel state side information at the transmitter/receiver by CSIT/CSIR.

at low SNR. We address this question by analyzing the effects of incomplete and/or noisy CSI.

With full CSI at both ends, the fading channel with i.i.d. fading or i.i.d. block fading is converted into a set of parallel memoryless channels [5], each corresponding to a certain value of the state $\mathbf{h}$. 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 without any *a priori* CSI. Consider the block fading Rayleigh channel with feedback[5] with no *a priori* CSI. If the coherence interval T is large enough, a negligible fraction of the total energy and time in each block can be spent in training and essentially perfect channel state estimates can be obtained at the receiver in each block. The same estimates also become available at the transmitter due to feedback. Thus, the capacity with full CSI can be achieved.

However, for smaller values of T, obtaining such estimates gets difficult due to low SNR. Even using all the available energy in a block for training may not be sufficient for channel state estimation. For example in the extreme case of $T = 1$, no CSI can be obtained at all. The value of feedback reduces monotonically with T. We aim to understand this effect more precisely-for this purpose we consider the joint asymptote of T and SNR, where T grows to infinity as SNR vanishes to zero. Section IV concludes with some engineering guidelines.

A. Summary of Results

This paper focuses on the low SNR limit for this fading channel. We use the notation $f(\mathsf{SNR}) \approx g(\mathsf{SNR})$ as a shorthand[6] for

$$\lim_{\mathsf{SNR} \to 0} \frac{f(\mathsf{SNR})}{g(\mathsf{SNR})} = 1.$$

Note that this is tighter definition of $\approx$ than $\lim_{\mathsf{SNR} \to 0} f(\mathsf{SNR}) = \lim_{\mathsf{SNR} \to 0} g(\mathsf{SNR})$ for approximating $f(\mathsf{SNR})$ with $g(\mathsf{SNR})$. Such a definition is too loose an approximation at low SNR, because most limits tend to zero anyway with SNR.

- Next section discusses the capacity of this channel without feedback in different cases of available CSI at the transmitter/receiver. With full CSI at the transmitter and receiver, let $C_{TR}(\mathsf{SNR})$ denote the capacity of the Rayleigh fading channel

$$C_{TR}(\mathsf{SNR}) \approx \mathsf{SNR} \cdot \log \frac{1}{\mathsf{SNR}}.$$

  Moreover, such rate is achievable even with a single bit of CSIT and no CSIR.
- We then address the effect of noisy CSI, where the actual channel state $\mathbf{h}$ is a noisier version of the channel state estimate $\mathbf{g}$ at the transmitter and receiver:

$$\mathbf{h} = \mathbf{g} + \mathbf{f},$$

---

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

[6]Inequalities, $f(\mathsf{SNR}) \overset{\sim}{>} g(\mathsf{SNR})$ and $f(\mathsf{SNR}) \overset{\sim}{<} g(\mathsf{SNR})$ are similarly defined.

where $\mathbf{g} \sim \mathcal{CN}(0, \beta)$ is the channel state estimate and $\mathbf{f} \sim \mathcal{CN}(0, 1 - \beta)$ is the estimation noise which independent of $\mathbf{g}$. The channel capacity in this case be denoted by $C_\beta(\mathsf{SNR})$

$$C_\beta(\mathsf{SNR}) \approx \beta \left(\mathsf{SNR} \cdot \log \frac{1}{\mathsf{SNR}} \right).$$

  Again, only one bit information of $\mathbf{g}$ at the transmitter is enough to achieve such a rate without any receiver CSI.
- Using the results in first two sections, the third section considers a wideband fading channel with feedback but no *a priori* CSI. In the joint asymptote where T grows to infinity as SNR goes to 0, a rate $R(\mathsf{SNR})$ as follows can be achieved:

$$R(\mathsf{SNR}) \approx \min\{\mathsf{SNR} \log T, \mathsf{SNR} \log 1/\mathsf{SNR}\}.$$

  If the only utility of feedback is for power control, a simple genie argument is used to show that no better rate is achievable. Even without any such assumption on the utility of feedback, $\mathsf{SNR} \log T$ is conjectured to be the feedback capacity.

II. VALUE OF THE CHANNEL STATE INFORMATION

This section considers the case where channel state $\mathbf{h}$ changes i.i.d. for every symbol. For the full CSI case, the energy efficiency, *i.e.*, capacity per unit energy is known to go to infinity in the low SNR limit [6]. Using properties of the Rayleigh distribution, we first prove that it goes to infinity as $\log(1/\mathsf{SNR})$ when the SNR goes to zero. Later we show how a single bit of CSIT can achieve the same energy efficiency. In Section II-B, we move to noisy CSI and show how the energy efficiency decreases with the level of noise in CSI.

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

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

Proof: For a channel strength of $t \triangleq |\mathbf{h}|^2$, the water-filling solution puts $p(t) \triangleq \left(\frac{1}{\Theta} - \frac{1}{t} \right)^+$ amount of power[7], where Θ is the channel gain threshold chosen to satisfy the average power constraint

$$\mathsf{SNR} = \int_0^\infty \left(\frac{1}{\Theta} - \frac{1}{t} \right)^+ e^{-t} dt = \int_\Theta^\infty \left(\frac{1}{\Theta} - \frac{1}{t} \right) e^{-t} dt \tag{4}$$

because t has an exponentional distribution of mean 1. Since a channel state of strength $t \geq \Theta$ contributes $\log(1 + tp(t)) = \log(t/\Theta)$ to capacity, the overall channel capacity $C_{TR}(\mathsf{SNR})$ in terms of the channel strength threshold Θ is given by $C_{\mathrm{thrd}}(\cdot)$ defined as follows:

$$C_{\mathrm{thrd}}(\Theta) \triangleq \int_\Theta^\infty \log(\frac{t}{\Theta}) e^{-t} dt. \tag{5}$$

---

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

We first show that the channel strength threshold α satisfying (4) 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 (4) can be written as $P_{\mathrm{thrd}}(\Theta)$, where $P_{\mathrm{thrd}}(\cdot)$ is defined as

$$P_{\mathrm{thrd}}(a) \triangleq \int_a^\infty \left(\frac{1}{a} - \frac{1}{t}\right) e^{-t}dt \qquad (6)$$

$$= \int_a^\infty \frac{e^{-t}}{t^2}dt \quad \text{(by integration by parts).} \quad (7)$$

Note that for all $a > 0$, this power lies between $P_L(a)$ and $P_U(a)$ defined as follows:

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

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

$$\lim_{\mathsf{SNR}\to 0} \frac{P_U(a_U)}{\mathsf{SNR}} = \frac{1}{1+\mu}$$

$$\text{Hence} \quad \lim_{\mathsf{SNR}\to 0} \frac{P_{\mathrm{thrd}}(a_U)}{\mathsf{SNR}} \le \frac{1}{1+\mu}.$$

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

$$\lim_{\mathsf{SNR}\to 0} \frac{P_L(a_L)}{\mathsf{SNR}} = 1+\mu$$

$$\text{Hence} \quad \lim_{\mathsf{SNR}\to 0} \frac{P_{\mathrm{thrd}}(a_L)}{\mathsf{SNR}} \ge 1+\mu.$$

Considering that $P_{\mathrm{thrd}}(a)$ is a strictly decreasing function, Θ 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 Θ 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 $C_{\mathrm{thrd}}(a)$ is a decreasing function and $\frac{e^{-a}}{a} - \frac{2e^{-a}}{a^2} \le C_{\mathrm{thrd}}(a) \le \frac{e^{-a}}{a}$ gives

$$\frac{e^{-a_U}}{a_U} - \frac{2e^{-a_U}}{a_U^2} \le C_{\mathrm{thrd}}(a_U) \le C_{\mathrm{thrd}}(\Theta) \le C_{\mathrm{thrd}}(a_L) \le \frac{e^{-a_L}}{a_L}.$$

We can verify that

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

$$\lim_{\mathsf{SNR}\to 0} \frac{e^{-a_U}/a_U - 2e^{-a_U}/a_U^2}{\log(1/\mathsf{SNR})\mathsf{SNR}} = \frac{1}{1+\mu}$$

$$\text{Hence} \quad \frac{1}{1+\mu} \le \lim_{\mathsf{SNR}\to 0} \frac{C_{\mathrm{thrd}}(\Theta)}{\log(1/\mathsf{SNR})\mathsf{SNR}} \le 1+\mu.$$

The lemma follows as arbitrarily small μ 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 On-Off scheme allocates a constant nonzero power level for any channel strength $|\mathbf{h}|^2 \ge \theta$ and zero power otherwise. This threshold θ for channel strength is equal to $\log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}})$.

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. It transmits at a nonzero power level SNR' for all $t \triangleq |\mathbf{h}|^2 \ge \log(\frac{1}{\mathsf{SNR}}) - 2\log\log(\frac{1}{\mathsf{SNR}}) \triangleq \theta$ and keeps silent otherwise. Since t is exponentially distributed with mean 1, the nonzero power level $\mathsf{SNR}' = \mathsf{SNR}/e^{-\theta}$. The achievable rate with this strategy equals

$$R(\mathsf{SNR}) = \int_\theta^\infty \log\left(1 + \frac{t\mathsf{SNR}}{e^{-\theta}}\right) e^{-t}dt$$

$$\ge \int_\theta^\infty \log\left(1 + \frac{\theta\mathsf{SNR}}{e^{-\theta}}\right) e^{-t}dt$$

$$= e^{-\theta}\log(1 + e^\theta\theta\,\mathsf{SNR}).$$

Note that $e^\theta\theta\mathsf{SNR}$ goes to zero as with SNR. Hence, $R(\mathsf{SNR}) \approx e^{-\theta} \cdot e^\theta\theta\,\mathsf{SNR} = \theta\mathsf{SNR}$. It is 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 almost always the received SNR in the above scheme, given by $t\mathsf{SNR}$, goes to zero with SNR. This explains the capacity achieving nature of the strategy. The channel state is almost always in the linear region, where the rate and received energy have a linear relationship. This gives an easy guiding principle for the low SNR asymptote of our interest: Stay in the linear region as much as possible.

The above power allocation only needed one bit of CSI at the transmitter for each fading realization. This bit indicated whether or not the channel strength is better than the threshold θ. Thus at low SNR, the capacity with only one bit of CSI at the transmitter is same as that with full CSI at the transmitter. This case is the "limited feedback" channel studied in [7], where the receiver has full CSI and the transmitter has its quantized version. Thus at low SNR, capacity of the limited feedback channel equals the capacity with full CSI at the transmitter.

Remark 1: Lemma 1 and 2 hold true even when no CSI is available at the receiver. [10] proves this by an orthogonal coding scheme where the receiver employs energy detection for decoding and energy is transmitted only where fading is sufficiently large[8].

Remark 2: For the channels in Lemma 1 and 2, [6] had shown previously that the ratio of the capacity and SNR goes to infinity with or without any receiver CSI. This behavior is described more precisely by Lemma 1: its goes to infinity as $\log(1/\mathsf{SNR})$.

---

[8]More specifically, Theorem 1 in [10] considers a Rayleigh fading wideband channel with causal transmitted CSI and no receiver CSI. It has a bandwidth of W symbols per unit time where each symbol i goes through the fading channel $\mathbf{y}_i = \mathbf{h}_i\mathbf{x}_i + \mathbf{w}_i$, where $\mathbf{h}_i$ and $\mathbf{w}_i$ are i.i.d. $\mathcal{CN}(0,1)$. If the transmitter has unit power available per unit time, the capacity of in the limit of large bandwidth W equals $C \approx \log W$ nats per unit time and is unchanged if even the receiver has full CSI. Thus, the capacity per unit symbol is $\approx \frac{\log W}{W}$ nats. Since $1/W$ denotes the SNR per symbol, this capacity is the same as $\mathsf{SNR}\log(1/\mathsf{SNR})$ in Lemmas 1 and 2.

A. 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 impossible to obtain such noiseless CSI. For example, if a channel is trained with energy E for obtaining CSI, the channel state estimate is noisy for any finite training energy $E > 0$. The actual channel state realization $\mathbf{h}$ in this case is given by the sum of two components: the channel state estimate $\mathbf{g}$ and the estimation error $\mathbf{f}$. The channel equation in this case is given by

$$\mathbf{y} = (\mathbf{g} + \mathbf{f})\mathbf{x} + \mathbf{w}. \tag{8}$$

We assume the channel state estimate $\mathbf{g} \sim \mathcal{CN}(0, \beta)$ and the error $\mathbf{f}$ is an independent $\mathcal{CN}(0, 1 - \beta)$. This model of independent complex Gaussian known part $\mathbf{g}$ and unknown part $\mathbf{f}$ is motivated by the way channel estimation is done in practice: by sending training signals. For Rayleigh channel $\mathbf{h}$, the MMSE channel estimate obtained form such training signal (i.e., $\mathbf{g}$) and the corresponding estimation error (i.e., $\mathbf{f}$) are independent of each other. This has motivated the channel model in (8) above. Note that when a training signal of energy E is used, the MMSE error variance for the channel state estimate equals $1/(1 + E)$ which is the same $1 - \beta$ in the channel model above. Thus, the channel model in (8) corresponds to a training signal with energy $E = \frac{\beta}{1-\beta}$.

We know that $\beta = 1$ corresponds to the full CSI case and $\beta = 0$ is no CSI case. At low SNR, their channel capacities are essentially $\log(1/\mathsf{SNR})\mathsf{SNR}$ and SNR, respectively. We expect intuitively that the capacity should increase with β. The following theorem shows how exactly it increases with β at low SNR. It considers β between the two extremes cases of $\beta = 0$ and $\beta = 1$ and shows that the noisy channel state estimate essentially reduces the capacity by a factor β.

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

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

$$C_\beta(\mathsf{SNR}) \approx \beta \cdot \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}.$$

Before going to the detailed proof, we sketch some intuition behind it. The capacity for channel in (8) with input power SNR can be upper bounded by the sum capacity of two separate channels-one with perfect transmitter CSI and one with no transmitter CSI. The input power available for each of these two channels is SNR (so total input power is $2\mathsf{SNR}$). The first channel $\mathbf{y}_1 = \mathbf{g}\mathbf{x}_1 + \mathbf{w}_1$, where $\mathbf{g} \sim \mathcal{CN}(0, \beta)$ is perfectly known at the transmitter and $\mathbf{w}_1 \sim \mathcal{CN}(0, 1)$. The other channel is $\mathbf{y}_2 = \mathbf{f}\mathbf{x}_2 + \mathbf{w}_2$, where the transmitter has no knowledge of $\mathbf{g} \sim \mathcal{CN}(0, \beta')$ and $\mathbf{w}_2 \sim \mathcal{CN}(0, 1)$. As we have seen already, the capacity of the first channel is $\approx \beta\mathsf{SNR}\log(1/\mathsf{SNR})$ and

that of the second channel is $\approx \beta'\mathsf{SNR}$. Hence the capacity of the original channel is at most $\approx \beta\mathsf{SNR}\log(1/\mathsf{SNR})+\beta'\mathsf{SNR} \approx \beta\mathsf{SNR}\log(1/\mathsf{SNR})$.

For achieving this upper bound, the transmitter simply ignores the unknown part of fading $\mathbf{f}$ and pretends that the actual channel $\mathbf{h}$ is equal to its known part $\mathbf{g}$. Then it applies essentially the same waterfilling strategy for the case of full transmitter CSI based on $\mathbf{g}$.

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

$$C_\beta(\mathsf{SNR}) = \max_{P(\mathbf{g}):\mathsf{E}[\mathbf{x}^2]\leq\mathsf{SNR}} \mathsf{E}\left[\log(1 + |\mathbf{g} + \mathbf{f}|^2 P(\mathbf{g}))\right]$$

because the transmitter only knows $\mathbf{g}$, which is independent of $\mathbf{f}$, the transmit power is a function of $\mathbf{g}$ only, so let us denote it by $P(\mathbf{g})$. The maximization above is over all such power allocation functions $P(\mathbf{g})$ satisfying the input power constraint. Let $P^*(\mathbf{g})$ denote the optimum choice of $P(\mathbf{g})$. Now Jensen's inequality and the independence between $\mathbf{g}$ and $\mathbf{f}$ implies

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

The first term above is the capacity of AWGN channel with signal to noise ratio of $\beta'\mathsf{E}\left[P^*(\mathbf{g})\right] = \beta'\mathsf{SNR}$. The second term is the capacity of a Rayleigh fading channel with transmit power SNR

$$\mathbf{y} = \mathbf{g}\mathbf{x} + \mathbf{w} \quad \text{where } \mathbf{g} \sim \mathcal{CN}(0, \beta) \tag{9}$$

where $\mathbf{g}$ is fully known at both ends. The optimal $P^*(\mathbf{g})$ is given by the water-filling solution. The above channel can be converted to the standard unit variance Rayleigh fading channel by dividing both sides of (9) by $\sqrt{\beta}$. This increases the noise power to $1/\beta$ and hence reduces the SNR to $\beta\mathsf{SNR}$. Now applying Lemma 1 gives

$$\lim_{\mathsf{SNR}\to 0} \frac{\mathsf{E}\left[\log\left(1 + |\mathbf{g}|^2 P^*(\mathbf{g})\right)\right]}{\log\left(\frac{1}{\beta\mathsf{SNR}}\right) \cdot \beta\mathsf{SNR}} \leq 1. \tag{10}$$

Hence, we can write

$$\frac{C_\beta(\mathsf{SNR})}{\beta'\mathsf{SNR} + \beta \cdot \log\left(\frac{1}{\beta\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1. \tag{11}$$

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

$$\frac{C_\beta(\mathsf{SNR})}{\mathsf{SNR} + \beta \cdot \log\left(\frac{1}{\beta\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1 \tag{12}$$

$$\Rightarrow \lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\mathsf{SNR} + \beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1 \tag{13}$$

$$\text{and} \quad \lim_{\mathsf{SNR}\to 0} \frac{C_\beta(\mathsf{SNR})}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} \leq 1. \tag{14}$$
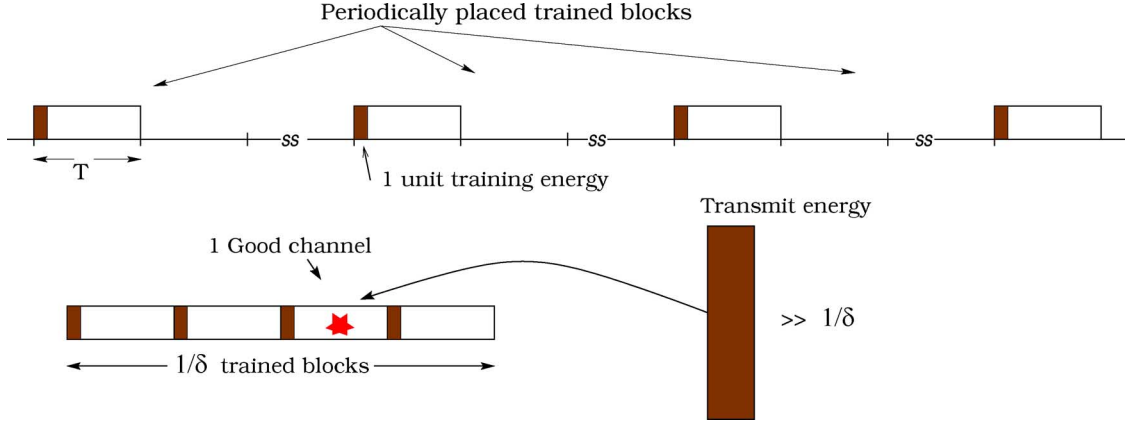
Fig. 1.   Training scheme for achievability.

For the lower bound to the capacity, we give an achievable scheme as follows. Transmit uniform power when the known channel's strength exceeds a threshold Θ_β defined as follows:

$$t_g \triangleq |\mathbf{g}|^2 \geq \beta \log\left(\frac{1}{\mathsf{SNR}}\right) - 2 \log\log\left(\frac{1}{\mathsf{SNR}}\right) \triangleq \Theta_\beta. \quad (15)$$

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

$R(\mathsf{SNR})$
$$\geq \mathsf{E}\left[\log\left(1 + |\mathbf{g} + \mathbf{f}|^2 e^{\Theta_\beta/\beta}\mathsf{SNR}\right) \cdot \mathbf{1}_{|\mathbf{g}|^2 \geq \Theta_\beta}\big|\, \mathbf{f}\right]$$
$\quad (\mathbf{1}_{|\mathbf{g}|^2 \geq \Theta_\beta}$ denotes indicator function$)$
$$\geq \mathsf{E}\left[\log\left(1 + (|\mathbf{g}| - |\mathbf{f}|)^2 e^{\Theta_\beta/\beta}\mathsf{SNR}\right) \cdot \mathbf{1}_{|\mathbf{g}|^2 \geq \Theta_\beta}\big|\, \mathbf{f}\right]$$
$$\geq \mathsf{E}\left[\log\left(1 + (|\mathbf{g}| - A)^2 e^{\Theta_\beta/\beta}\mathsf{SNR}\right) \cdot \mathbf{1}_{|\mathbf{g}|^2 \geq \Theta_\beta} \cdot \mathbf{1}_{|\mathbf{f}| < A}\right]$$
$\quad$ (for some fixed $A > 0$)
$$= (1 - e^{A/\beta'}) \cdot e^{-\Theta_\beta/\beta} \cdot \log\left(1 + |\sqrt{\Theta_\beta} - A|^2 e^{\Theta_\beta/\beta}\mathsf{SNR}\right).$$

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 in (15). Hence

$$\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \cdot \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}} = \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\Theta_\beta \mathsf{SNR}}$$
$$\geq \lim_{\mathsf{SNR}\to 0} (1 - e^{A/\beta'}) \cdot \left|1 - \frac{A}{\sqrt{\Theta_\beta}}\right|^2$$
$$= (1 - e^{A/\beta'}) \cdot 1.$$

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 that the threshold Θ_β is essentially β times the threshold for the full CSI case. Hence, the probability $\exp(-\Theta_\beta/\beta)$ of transmitting nonzero power is the same as that in the full CSI case. Thu,s although the fraction of channel states used is same as the full CSI case, the noise in the transmitter CSI reduces the capacity by a factor of β.

III. FEEDBACK CHANNEL WITH 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. We study how exactly it increases with T in the limit of large T. More precisely, we assume $T \to \infty$ as $\mathsf{SNR} \to 0$, for example $T(\mathsf{SNR}) = \log(1/\mathsf{SNR})$. One special case of this assumption is $T(\mathsf{SNR}) = \mathsf{SNR}^{-d}$ as in [11].

Theorem 4: For $T \leq 1/\mathsf{SNR}$, as $\mathsf{SNR} \to 0$, a data rate $R(\mathsf{SNR})$ is achievable such that

$$\frac{R(\mathsf{SNR})}{\mathsf{SNR} \log T} \to 1.$$

Proof: We train one out of every $1/\delta_1 = E^2/T\mathsf{SNR}$ fading blocks, where E denotes a large but fixed the training energy. This periodic placement of trained blocks (1 after $1/\delta_1$ fading blocks) is illustrated in the top picture in Fig. 1. The training signal is sent on the first symbol of the training block (the shaded strip in Fig. 1 at the beginning of each training block). The remaining $T - 1$ symbols in each training block are used for communication only. The positions of these periodically spaced training blocks are predetermined and conveyed to both the ends. The transmitter is silent ($\mathbf{x} = 0$) throughout any nontraining block.

For each trained block, the total energy accumulated over $1/\delta_1$ blocks equals $E_{\text{Total}} = T\mathsf{SNR} \cdot \frac{E^2}{T\mathsf{SNR}} = E^2$. This is because the transmitter is silent throughout any untrained fading block. Since the training signal has energy E, the fraction of this total energy used for the training signal equals E/E^2 and can be ignored as E is large. Hence, the average SNR available in the trained blocks equals

$$\mathsf{SNR}' = E_{\text{Total}}/T = E^2/T.$$

After using the first symbol in each training block for training, for the remaining $T - 1$ symbols in that block, we get the channel

with partial CSI in Theorem 3. Hence, we can apply the communication scheme in Theorem 3 over all the trained blocks. Recall that a training energy of E corresponds to $\beta = \frac{E}{1+E}$ in the previous section. 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, recall that only $\delta_1 = T\mathsf{SNR}/E^2$ fraction of the blocks are trained. Moreover, in each training block of length T, only $T-1$ symbols are used for communication (since its first symbol is reserved for the training signal). Hence, the overall rate achieved with this scheme satisfies

$$\lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\delta_1 \cdot \beta \mathsf{SNR}' \log(1/\mathsf{SNR}')} = \lim_{\mathsf{SNR}\to 0} \frac{T-1}{T} = 1$$
$$(\text{since } T\to\infty \text{ as } \mathsf{SNR}\to 0)$$
$$\Rightarrow \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \mathsf{SNR} \log(T/E^2)} = \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \mathsf{SNR} \log(T/E^2)}$$
$$(\text{since } \delta_1 \mathsf{SNR}' = \mathsf{SNR})$$
$$= \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\beta \mathsf{SNR} \log T}$$
$$= 1$$
$$\Rightarrow \lim_{\mathsf{SNR}\to 0} \frac{R(\mathsf{SNR})}{\mathsf{SNR} \log T} = \beta = \frac{E}{1+E}.$$

Choosing arbitrarily large training energy E yields the proof.
∎

To interpret this result, let us assume that (roughly) unit energy is needed to train a channel block perfectly. This is an optimistic assumption because perfect CSI cannot be obtained with finite training energy, however such finite training energy can achieve sufficiently accurate CSI. Also, assume that 1 out of every $1/\delta$ trained blocks is "good", i.e., above the channel threshold. This channel threshold should be $\log(1/\delta)$ due to exponential nature of Rayleigh distribution. Note that $1/\delta$ training energy should be spent for finding one good channel state. 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. Since the transmit power is ratio of the received power ρ to the channel gain (essentially $\log(1/\delta)$), 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 state (see Fig. 1), the energy $\frac{\rho T}{\log(1/\delta)}$ transmitted in the good channel state should be much larger than $1/\delta$. Thus, δ should be larger than $1/T$. Hence the channel strength threshold $\log(1/\delta)$, i.e., the energy efficiency cannot be larger than $\log T$.

Note that very few bits of feedback are needed on average in this achievability proof. Feedback is only needed for the trained channel blocks. Moreover, as seen in the previous section, only one bit of feedback (per fading realization) about the trained channel block is good enough for achieving its capacity at low SNR. This bit should indicate whether or not the channel state estimate is better than the threshold Θ_β. Thus, on average, a feedback of δ_1 bits per block or $\delta_1/T = \mathsf{SNR}/E^2$ bits per symbol is required. Thus a very weak (but reliable) reverse channel is sufficient for this achievable strategy.

Remark 4: 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 $|h|^2_{\max}$. For any coherence interval T going 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 $C(\mathsf{SNR}) \approx |h|^2_{\max}\mathsf{SNR}$.

For $T \gg 1/\mathsf{SNR}$, the energy per coherent block $T\mathsf{SNR} \gg 1$. Hence, essentially perfect CSI can be obtained by spending a negligible fraction this energy on training each block. Hence the capacity with full CSI at both ends (i.e., essentially a rate of $\mathsf{SNR}\log(1/\mathsf{SNR})$) is achievable by training. The above theorem shows that for any $T \geq 1/\mathsf{SNR}$ (not only $T \gg 1/\mathsf{SNR}$), full CSI capacity is achievable by the proposed peaky training based scheme.

Since the capacity with full CSI at both ends is an upper bound to the capacity of this feedback channel, the strategy for Theorem 4 achieves the capacity of this feedback channel when $T \geq \mathsf{SNR}^{-1}$. For smaller coherence interval, we need a tighter upper bound than the capacity with full CSI at both ends. For achieving the feedback capacity, assuming that the transmitter only needs the feedback to obtain CSI at the transmitter and correspondingly adjust the transmitted power, the following theorem gives an upper bound to the achievable rate. First, let us define precisely the notion of using feedback only for power control.

Definition 1: Power-Control-Only feedback codes: This is a class of feedback codes which use feedback only to obtain CSI at the transmitter and correspondingly adjust the transmitted power. Mathematically, these codes satisfy the following Markov condition $(m, \mathbf{x}_1^t, \mathbf{y}_1^t) - (m, \hat{\mathbf{h}}_{t+1}) - \mathbf{x}_{t+1}$, where m denotes the message to be transmitted and $\hat{\mathbf{h}}_t + 1$ denotes the transmitter's estimate of the channel state $\mathbf{h}_{t+1}$. This estimate is based on $(\mathbf{x}_1^t, \mathbf{y}_1^t)$. Thus the only influence of past inputs and outputs on the next input $\mathbf{x}_t$ is through the channel state estimate $\hat{\mathbf{h}}_t$.

For *Power-Control-Only* feedback codes, the upper bound in the next theorem matches with the achievable rate when T grows to infinity with SNR.

Theorem 5: For *Power-Control-Only* feedback codes, the capacity of the block fading Rayleigh channel with feedback satisfies

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

for any T that goes to infinity as SNR tends to 0 such that $T \leq 1/\mathsf{SNR}$.

Proof: The achievability part of this capacity result was already proved in Theorem 4, so we only need to prove the converse part now. We show that achievable rate for any *Power-Control-Only* feedback code is upper bounded as

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

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

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.

The genie also provides us free additional transmit power of SNR for this channel. Since the quality of the channel state 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 communications channel just before its corresponding block starts.

Another genie tells us the exact channel state $\mathbf{h}$ 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 (13), the capacity of a Rayleigh channel with $(1 - \beta)$ estimation error at the transmitter is upper bounded by $\mathsf{SNR} + \beta \log(\frac{1}{\mathsf{SNR}})\mathsf{SNR}$ in the limit of low SNR. Now consider a distribution of training energies on the training channel. We assume this to be a discrete distribution for simplicity, but generalization to continuous distributions is not difficult. Say training with energy E_j is performed with probability p_j and the communication SNR for this training level equals SNR_j. Assume that nonperfect 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

$$p_i R_i(\mathsf{SNR}_i) \overset{\sim}{<} p_i \cdot \left(\mathsf{SNR}_i + \frac{E_i}{1+E_i} \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i \right) \quad (16)$$

$$\leq p_i \cdot \left(\mathsf{SNR}_i + E_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i \right) \quad (17)$$

$$\approx p_i E_i (\mathsf{SNR}_i \log(\frac{1}{\mathsf{SNR}_i})). \quad (18)$$

The last step follows because SNR_i is negligible compared to $\mathsf{SNR}_i \log(\frac{1}{\mathsf{SNR}_i})$ in the limit of low SNR. The above inequality implies the training with $E_i < 1$ with probability p_i achieves a rate (given by $p_i R_i(\mathsf{SNR}_i)$) worse than training perfectly with $E = 1$ with probability $p_i E_i$ [which achieves a rate $\approx p_i E_i(\mathsf{SNR}_i \log(1/\mathsf{SNR}_i))$]. The total training energy for training with probability p_i at energy E_i is $p_i E_i$, which is the same as the energy required for training with probability $p_i E_i$ with $E = 1$. 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 (17) as

$$\sum_i p_i R_i(\mathsf{SNR}_i)$$

$$\leq p_0 \mathsf{SNR}_0 + \sum_{i \neq 0} p_i \cdot \left(\mathsf{SNR}_i + E_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i \right)$$

$$\leq \sum_i p_i \mathsf{SNR}_i + \sum_{i \neq 0} p_i \log(\frac{1}{\mathsf{SNR}_i})\mathsf{SNR}_i \text{ because all } E_i \leq 1.$$

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

$$\mathrm{RHS} \leq \mathsf{SNR} + \mathsf{SNR} \log(\frac{1 - p_0}{\mathsf{SNR}})$$

$$= \mathsf{SNR} + \mathsf{SNR} \log(\frac{p_1}{\mathsf{SNR}}) \text{ where } p_1 = 1 - p_0.$$

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 on this genie-aided channel. 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 above upper bound is increasing in p_1, we chose the maximum possible p_1. Hence, the capacity of this channel satisfies

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

$$\bullet$$

Note that the total power used for this proof is twice the available power. The training channel uses SNR units of power and communication over trained channel blocks uses another SNR units.

We conjecture that even if the utility of feedback is not restricted to power control, the same upper bound of $C(\mathsf{SNR}) \overset{\sim}{<} \mathsf{SNR} \log T$ holds for general feedback strategies. Since this upper bound is already shown to be achievable, this conjecture is equivalent to claiming that the feedback capacity of this block fading channel is $\approx \mathsf{SNR} \log T$. The key part missing in proving this conjecture is showing that for any capacity achieving code, peakiness[9] of the received power should increase monotonically as the coherence interval (and hence the ability to estimate the channel) increases.

IV. CONCLUDING REMARKS

At low SNR, the engineering rules learned for block fading channels with feedback are as follows.

1) The received SNR should be small as often as possible to take advantage of the linear rate-power behavior.
2) Whenever training is performed, it 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 take over the trained blocks.

It is worth mentioning that similar analysis can be applied even when the fading distribution is not Rayleigh. Results in those cases will depend on the tail behaviour of the fading distribution. For example, let the tail of the fading gain distribution be a polynomial, that is, probability $\Pr(|\mathbf{h}|^2 \geq \mathsf{x})$ behaves as x^{-n} for a given $n > 0$. Using results in [10], the feedback capacity in this case is expected to be essentially $T^{\frac{1}{n+1}}\mathsf{SNR}$ nats/symbol.

Second, we remark on a one step Gauss-Markov fading with innovation rate υ going to zero with SNR. This can be converted to a block fading of essentially a coherence interval of $1/\upsilon$. Such conversion ignores the correlation between adjacent blocks. By Theorem 4, a rate of $\mathsf{SNR} \log(1/\upsilon)$ can be achieved at low SNR. However, optimality of such conversion to block fading is not known.

Finally, this setup can be used for the wideband broadcast fading channel with a feedback link from each receiver. Here, a single training pulse is enough to estimate all the users'

---

[9]We define peakiness of the received power as the ratio of its second moment to first moment.

channels. Assuming the various users are independently faded, the probability of finding at least one user with strong channel state is increased. Thus, training is more beneficial than before. For the case of a fixed number of users, the sum feedback capacity is not expected to change much from the single user case. Nonetheless, for the case of a large number of users (which grows to infinity as SNR vanishes), the broadcast nature of training pulses can improve the sum capacity significantly.

ACKNOWLEDGMENT

We are thankful to E. Telatar for his help for his many insightful comments.

REFERENCES

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

[2] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[3] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity and zero-error capacity of the chemical channel with feedback," in *Proc. IEEE ISIT*, Nice, France, 2007.

[4] H. Permuter, T. Weissman, and A. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inform. Theory*, submitted for publication.

[5] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[6] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1342, Jun. 2002.

[7] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2003.

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

[9] S. Borade and L. Zheng, "Wideband fading channels with feedback," in *Proc. Allerton Conf.*, Monticello, IL, Sep. 2004.

[10] S. Borade and L. Zheng, "Writing on fading paper and causal transmitter CSI," in *Proc. IEEE Symp. Information Theory*, Seattle, WA, Jul. 2006.

[11] L. Zheng, D. N. C. Tse, and M. Medard, "Channel coherence in the low SNR regime," in *Proc. Allerton Confer.*, Monticello, IL, Oct. 2004.

Shashi Borade (S'01–M'10) received the B.Tech. degree in electrical engineering from Indian Institute of Technology, Mumbai, India, in 2002 and the S.M. and Ph.D. degrees in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge, in 2004 and 2008, respectively.

His research interests include statistics, information theory, and finance theory. Since 2008, he has been a Quantitative Researcher at D. E. Shaw and Co., New York.

Dr. Borade received the Institute Silver Medal from IIT Bombay in 2002, the Presidential Fellowship from MIT in 2002, and the Hewlett Packard Graduate Fellowship in 2005.

Lizhong Zheng (S'00–M'02) received the B.S. and M.S. degree from the Department of Electronic Engineering, Tsinghua University, Tsinghua, China, in 1994 and 1997, respectively, and the Ph.D. degree from the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, in 2002.

Since 2002, he has been with the Department of Electrical Engineering and Computer Sciences, Massachusetts Institute of Technology, Cambridge, where he is currently the KDD Associate Professor. His research interests include information theory, wireless communications , and wireless networks.

Dr. Zheng received the Eli Jury Award from the University of California, Berkeley, in 2002, the IEEE Information Theory Society Paper Award in 2003, and the National Science Foundation CAREER award in 2004. He is currently an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY.