On Channel Coherence in the Low SNR Regime *

Lizhong Zheng Muriel Médard David Tse Cheng Luo †

Abstract

The current study of the low SNR fading channels focuses on two extreme cases: the coherent case with perfect channel state information (CSI) available at the receiver, and the non-coherent case with no hope to obtain the channel information at all. Most of practical channels, with a low SNR and a slowly varying channel, lies in between these two extremes. In this paper, we use training based schemes to take the advantage of such channel coherence in the low SNR regime. We characterize how slowly the channel can change over time such that a "near coherent" performance to be achieved. We demonstrate that use training scheme in a flashy fashion can improve the performance. We also defined the notion of "operation coherence level", which is used to describe a continuum between the coherent and the non-coherent extremes.

1 Introduction

The study of using a large bandwidth to improve the power efficiency in wireless communications dates back to 1960's. Kennedy [1] showed that for a Rayleigh fading channel at the infinite bandwidth limit, the amount of energy required to reliably transmit one information bit is $\frac{E_b}{N_0} = -1.59dB$, which is the same limit for the AWGN channels. Denote the signal-to-noise ration per degree of freedom as SNR, and the corresponding capacity as $C(\mathsf{SNR})$, this result is equivalent as

$$\lim_{\mathsf{SNR}\to 0} \frac{C_{fading}(\mathsf{SNR})}{\mathsf{SNR}} = \lim_{\mathsf{SNR}} \frac{C_{AWGN}(\mathsf{SNR})}{\mathsf{SNR}} = 1 \tag{1}$$

This result is very robust. it applies to both the cases that the instantaneous channel state information is/is not available at the receiver, referred as the coherent/ noncoherent channels, respectively, in this paper. It is later shown [2] that (1) also holds for general fading distributions.

*The research of M. Médard is supported in part National Science Foundation, "CAREER", CCR-00993349 and Hewlett-Packard, "Advanced Concepts in Wireless Networking", 008542-008. The research of D. Tse is supported by a MURI grant on Ultra-Wideband Communications.

†Zheng, Médard, and Luo are with the department of EECS, Massachusetts Institute of Technology, David Tse is with the department of EECS at University of California, Berkeley

However, it is also observed [2, 5] that it is much harder to approach the wideband limit for the non-coherent channel than for the coherent channel. Using the standard spread spectrum techniques in the non-coherent channel cannot achieve the wideband capacity limit. In fact this limit can only be achieved by transmitting flashy signals, which means transmitter must concentrate all its power must in a vanishing fraction of the available time slots, and remain silent for the rest of the time. Such signaling, with large peak-to-average ratio, is hard to implement in practice. Moreover, it is also observed that the wideband limit is approached much slower for the non-coherent case than the coherent case as the bandwidth increases. Such important practical differences between the coherent and the non-coherent channels are not captured by the result (1).

Recently, Verdu addressed this issue quantitatively by looking at the second order Tyler expansion of the capacity expression:

$$C(\mathsf{SNR}) = C'(0)\mathsf{SNR} + C''(0)\mathsf{SNR}^2 + o(\mathsf{SNR}^2). \tag{2}$$

where $C'(0)$ and $C''(0)$ are the first and second derivatives of the function $\mathsf{SNR} \to C(\mathsf{SNR})$ at $\mathsf{SNR} = 0$. It is shown that for a general class of channels, when CSI is not available at the receiver, it is necessary that the transmitted signals to be "flashy", in order to achieve the first order optimal $C'(0) = 1$. However, with flash signals, the second derivative $C''(0)$ is always $-\infty$, in contrast to the capacity of a channel with perfect CSI at the receiver, for which $C''(0)$ is finite. Thus the difference between the coherent and the noncoherent cases lies in the sub-linear terms of the capacity expressions.

While the difference between the coherent and the noncoherent cases is clarified, the connections between the two cases are not well-understood. In practice, the coherent/noncoherent assumptions are made based on the speed that the channel changes over time. Intuitively, as the channel coherence time increases, one can afford a large amount of time and energy for channel estimation, without affecting the overall performance. Therefore, the near-coherent performance can be achieved even if the CSI is not available in the first place. The current notion of the "coherent channel" assumes perfect channel information, without covering the cases with partial channel information; on the other hand, the current notion of the "noncoherent channel" assumes the channel realizations to be independent from one symbol period to another, without taking the advantage of slow channel variation. These two cases, which will be continue referred as the coherent/noncoherent cases throughout this paper, can thus be viewed as two extreme cases. The performance gap between these two extremes is significant.

A natural question is therefore: how slowly the channel can change over time such that the capacity for a non-coherent channel starts to "look like" that for the coherent channel, or the gap between the two extremes being closed.

In this paper, we address this issue by studying the performance of training-based schemes, used in a non-coherent block fading channel. We give a sufficient condition on the channel coherence time in order that $C''(0)$ in (2) of the non-coherent channel to be finite. In other words, we characterize the coherence time required for a non-coherent block fading channel to behave close to a channel with perfect CSI at the receiver. Furthermore, we define the notion of "operational coherence level", in order to describe a continuum between the coherent and the non-coherent extremes. We also characterize the coherence time required to achieve an intermediate level of coherence, thus bridge the gap between these two extremes.

In addition to the conventional training schemes, we also developed the "flash training" schemes, which only send training signals and communicate over a small fraction of the available coherence blocks. Such schemes are particularly useful in the low SNR regime. We demonstrate the performance improvement by using such flash schemes, and characterize the optimal level of flashiness as a function of the channel coherence time and SNR.

2 Preliminaries and Problem Formulation

In this paper, we consider a block fading channel, which can be written as

$$\mathbf{y}_i = \sqrt{\mathsf{SNR}}\mathbf{h}\mathbf{x}_i + \mathbf{w}_i, \quad i = 1, \ldots, l$$

where $\mathbf{x}_i, \mathbf{y}_i$ are the transmitted and received signals, respectively, at time instance i, and $\mathbf{w}_i$ is the additive white Gaussian noise with unit variance. The fading coefficient $\mathbf{h}$ is assumed to be also $\mathcal{CN}(0, 1)$ distributed. More over we assume that $\mathbf{h}$ remains constant over a block of l symbols, and after that changes into independent realizations. The block length l, also referred as the *coherence time*, indicates the speed that the fading channel changes over time. The power constraint for this channel is normalized so that $E[|\mathbf{x}_i|^2] = 1$.

We consider a communication system over many of such independently faded coherence blocks, resulting from communicating over a long period of time and/or over a wide frequency band that covers multiple coherence bandwidth.

It is well known that at the low SNR regime, the capacity of this channel, whether the fading coefficients are known to the receiver or not, is close to a linear function of SNR, i.e.,

$$C(\mathsf{SNR}) = \mathsf{SNR} + o(\mathsf{SNR})(\text{nats/symbol}) \tag{3}$$

However, the sub-linear term $\mathsf{SNR} - C(\mathsf{SNR})$ can be very different depending on the availability of the channel state information at he receiver.

For the coherent case, i.e., when $\mathbf{h}$ is perfectly known at the receiver, the channel capacity at low SNR can be written as

$$C_{coherent}(\mathsf{SNR}) = \mathsf{SNR} - O(\mathsf{SNR}^2)$$

That is, the sub-linear term $\mathsf{SNR} - C(\mathsf{SNR})$ is of the order $O(\mathsf{SNR}^2)$. On the other hand, it is shown in [4] that for the non-coherent channel, for which $\mathbf{h}$ is not known, the sub-linear term in of the capacity is much larger,

$$\mathsf{SNR} - C_{non-coherent}(\mathsf{SNR}) \gg \mathsf{SNR}^2 \tag{4}$$

where $\gg$ means the ratio between the two sides goes to infinity at low SNR.

The gap between the coherent and the non-coherent cases are thus obvious in the sub-linear term of the capacity. Although this contribute to only a small fraction of the capacity, it is shown in [4] that this difference becomes important when communicating with a constraint on the energy per information bit. The bandwidth required to achieve a particular energy efficiency level is much larger for the non-coherent case than for the coherent case. In other words, the non-coherent channel converges to the wide band limit much slower.

Intuitively, if the channel changes slowly over time, then one can use a significant amount of energy in training to help the receiver to estimate the channel precisely, yet this cost of energy is only an ignorable fraction of the total energy. Thus as the coherence time increases, we should have a continuum between the coherent and the non-coherent extremes. The result in [4], however, does not describe such a continuum. It is stated that for any given coherence time l, as SNR approach 0, (4) always holds, which means that a near coherent performance (3) can never be achieved. In fact, whether a channel, with particular values of the coherence time and SNR, is closer to the coherent extreme or the non-coherent extreme depends on the relation between the coherence time and the SNR, instead either parameter alone. By fixing the coherence time and let the SNR go to 0, this key relation is distorted, thus we always get to the non-coherent extreme.

To address this problem, in this paper, we study the channel with both the SNR goes to 0 and the coherence time l goes to infinity, while maintaining a relation between the two, described as $l(\mathsf{SNR})$. We ask the question as "for what function $l(\mathsf{SNR})$, or how fast needs the coherence time goes to infinity, in order that a near coherent throughput (3) can be achieved?" We answer this question by studying the performance of a specific training based scheme.

To describe a continuum between the two extremes, we also studied the problem of how large the coherence time has to be, in order that the throughput behaves like

$$\mathsf{SNR} - O(\mathsf{SNR}^{1+\alpha})$$

for $\alpha \in [0, 1]$, where $\alpha = 0$ corresponds to the non-coherent extreme and $\alpha = 1$ corresponds to the coherent. Thus the intermediate values of α can be called the *operational coherence level* of the channel.

3 Training schemes for the Block fading model

In this section, we study a sub-optimal approach to use the block fading channel, namely, by using training schemes. Training schemes are widely used in communicating over fading channels when the fading coefficients are unknown to the receiver. At the beginning of each coherence block, a training sequence, known to the receiver, is transmitted to help the receiver to estimate the channel coefficients, and then these estimates are used to communicate during the rest of the coherence block. We follow the approach used in [3] and [6], to compute a lower bound of the achievable throughput by optimizing the amount of energy used in training.

We start by describing the training scheme in details. We rewrite the block fading channel model, within one coherence block, as follows

$$\mathbf{y}_i = \sqrt{\mathsf{SNR}}\mathbf{h}\mathbf{x}_i + \mathbf{w}_i, \quad i = 1, \ldots, l$$

where the fading coefficient $\mathbf{h}$ is assumed to remain constant within a block of l symbols. For convenience, we refer to $\sqrt{\mathsf{SNR}}\mathbf{x}_i$ as the "transmitted signal", which has an average energy per symbol time of SNR.

From the previous study, we observe that in order to maximize the throughput for a non-coherent channel at low SNR, a flashy input signal has to be used. That is, the transmitted energy needs to be concentrated in a small fraction of the available time slots. Intuitively, this is because that the available energy is not enough to estimate all the channel coefficients. However, most of the current study on the training schemes assumes that signals are transmitted in every coherence block. Such treatment implicitly labelled training schemes as more suitable for the cases when the SNR is high and the coherence time is long, or in other words, when the channel is nearly coherent. It is shown in [3] that such training schemes is optimal at high SNR in achieving the maximum number of degrees of freedom, and in [5] that at the achievable rate by such schemes decreases to 0 at the wideband limit. Thus to adopt training schemes in the cases with transmit energy being limited, we need to study "flash training schemes". That is, we introduce flashiness into training schemes by concentrating the transmitted energy in a small fraction δ of coherence blocks. Within each of these blocks, training and communication take places as before, while the other coherence blocks are simply ignored. In this section, we will demonstrate the performance improvement by using this simple idea.

In a long time period, we choose to transmit signals in δ fraction of available coherence blocks. In each block that signals are transmitted, the total average energy is given by

$$E_{total} = \frac{1}{\delta} l \mathsf{SNR}$$

At the beginning of a block, we use γ fraction of the total energy in training. For convenience, denote the energy used in training as

$$E_{tr} = \gamma E_{total} = \frac{1}{\delta} \gamma l \mathsf{SNR}$$

The receiver computes the minimum mean square estimate of the fading coefficient $\mathbf{h}$. Since the quality of this estimate depends only on the energy, instead of the time period, of the training signal, we assume in the following that the training signal is transmitted within 1 symbol period, i.e., the signal transmitted in the first symbol of a block is

$$\sqrt{\mathsf{SNR}}\mathbf{x}_1 = \sqrt{E_{tr}}$$

and the receiver signal is

$$\mathbf{y}_1 = \sqrt{E_{tr}}\mathbf{h} + \mathbf{w}_1$$

Use $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ to denote the minimum mean square estimate of $\mathbf{h}$ and the estimation error, respectively, we have

$$\begin{aligned} E[|\hat{\mathbf{h}}|^2] &= \frac{E_{tr}}{1+E_{tr}} \\ E[|\tilde{\mathbf{h}}|^2] &= \frac{1}{1+E_{tr}} \end{aligned}$$

For the rest $(l-1)$ symbols within the same block, we communicate by using an i.i.d. Gaussian random code with average power of $(1-\gamma)\frac{1}{\delta}\mathsf{SNR}$. The channel in this communication phase can be written as

$$\begin{aligned} \mathbf{y}_i &= \sqrt{\frac{(1-\gamma)}{\delta}\mathsf{SNR}} \cdot \hat{\mathbf{h}}\mathbf{x}_i + \sqrt{\frac{(1-\gamma)}{\delta}\mathsf{SNR}} \cdot \tilde{\mathbf{h}}\mathbf{x}_i + \mathbf{w}_i \\ &= \sqrt{\frac{(1-\gamma)}{\delta}\mathsf{SNR}} \cdot \hat{\mathbf{h}}\mathbf{x}_i + \mathbf{w}'_i \end{aligned} \tag{5}$$

for $i = 2, \ldots, l$, and $\mathbf{x}_i$ is normalized to have $E[|\mathbf{x}_i|^2] = 1$. Notice that the second term in (5), the extra noise due to the channel estimation error, is uncorrelated with the signal term, $\hat{\mathbf{h}}\mathbf{x}_i$. Thus to obtain a lower bound of the mutual information, we can replace it with the additive Gaussian noise with the same power. The overall noise $\mathbf{w}'_i$ is replaced by a Gaussian noise with variance

$$\begin{aligned} \sigma^2 &= 1 + \frac{(1-\gamma)}{\delta} E[|\tilde{\mathbf{h}}^2]\mathsf{SNR} \\ &= 1 + \frac{(1-\gamma)}{\delta}\frac{1}{1+E_{tr}}\mathsf{SNR} \end{aligned}$$

The resulting mutual information per symbol time is lower bounded by

$$I_{tr}(\mathsf{SNR}, \delta) \geq \delta \frac{l-1}{l} E\left[\log\left(1 + \frac{\frac{(1-\gamma)}{\delta}\mathsf{SNR}|\hat{\mathbf{h}}|^2}{\sigma^2}\right)\right] \tag{6}$$

where the factor $\delta(l-1)/l$ is due to the fact that one δ fraction of the time is used in communication, and that 1 symbol time out of a block is used in training.

3.1 The Performance of Non-Flash Training Schemes

With the performance lower bound of the training schemes established, now we optimize over the power allocation γ, and the following lemma gives a lower bound of the achievable data rate for the conventional non-flashy training schemes, corresponding to $\delta = 1$. To simplify the notation, we write the mutual information for $\delta = 1$ as

$$I_{tr}(\mathsf{SNR}) \triangleq I_{tr}(\mathsf{SNR}, \delta = 1)$$

Lemma 1 *Fix $\delta = 1$. If the coherence time satisfies*

$$l(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(1+2\alpha)}$$

for $\alpha \in [0, 1]$, then the training scheme described above achieves a data rate $I_{tr}(\mathsf{SNR})$ with

$$\mathsf{SNR} - I_{tr}(\mathsf{SNR}) \dot{\leq} \mathsf{SNR}^{1+\alpha}$$

In other words, we can achieve a data rate of the order

$$I_{tr}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR} - O(\mathsf{SNR}^{1+\alpha})$$

Proof:
Omitted for this version of the paper.

This Lemma says that the coherence time required for a training schemes to achieve a certain coherence level α is $l \doteq \mathsf{SNR}^{-(1+2\alpha)}$, which is much larger than that for the channel capacity ($\mathsf{SNR}^{-2\alpha}$). In particular, if we wish to achieve a near coherent performance (3) with the training scheme, corresponding to the case $\alpha = 1$, Lemma 1 suggests that it is sufficient to have

$$l \doteq \mathsf{SNR}^{-3} \tag{7}$$

While Lemma 1 is only an achievable performance, in the following, we argue in the following that the condition (7) is indeed necessary for a training scheme, to achieve the near coherent performance in (3). Intuitively, if we want a training scheme for a non-coherent channel to have a throughput close to the capacity of the perfectly coherent channel, it is necessary that both the following conditions be satisfied:

- The energy used in training, E_{tr}, is large enough such that the channel estimation error is ignorable.
- The fraction of energy used in training, γ is small enough such that its effect is ignorable.

In the scaling of interests, these two conditions can be quantitatively specified.

For convenience, we write the RHS of (6) with $\delta = 1$ as [1]

$$R_{tr}(\mathsf{SNR}) = \frac{l-1}{l} E\left[\log\left(1 + \frac{(1-\gamma)\mathsf{SNR}|\hat{\mathbf{h}}|^2}{\sigma^2}\right)\right]$$

[1]Strictly speaking, $R_{tr}(\mathsf{SNR})$ is a lower bound of the achievable rate for the training schemes. However, in order to achieve a rate higher than $R_{tr}(\mathsf{SNR})$, a receiver that can take the advantage of the estimation error term, $\tilde{\mathbf{h}}\mathbf{x}$, is required, which is usually difficult. Therefore, we take $R_{tr}(\mathsf{SNR})$ as a "practical" approximation of $I_{tr}(\mathsf{SNR})$.

Notice that $\sigma^2 \geq 1$, and use Jansen's inequality, a simple upper bound of $R_{tr}(\mathsf{SNR})$ is given by

$$\begin{aligned} R_{tr}(\mathsf{SNR}) &\leq \log(1 + \mathsf{SNR}E[|\hat{\mathbf{h}}|^2]) \\ &\leq \mathsf{SNR}\frac{E_{tr}}{1+E_{tr}} \\ &= \mathsf{SNR} - \mathsf{SNR}\frac{1}{1+E_{tr}} \end{aligned}$$

In order that $R_{tr}(\mathsf{SNR}) \doteq \mathsf{SNR} - O(\mathsf{SNR}^2)$, it requires that the training energy to be as large as:

$$E_{tr} \dot{\geq} \mathsf{SNR}^{-1} \tag{8}$$

Another upper bound of $R_{tr}(\mathsf{SNR})$, noticing $E[|\hat{\mathbf{h}}|^2] \leq 1$, can be written as

$$\begin{aligned} R_{tr}(\mathsf{SNR}) &\leq \log(1 + (1-\gamma)\mathsf{SNR}) &(9) \\ &\leq \mathsf{SNR} - \gamma\mathsf{SNR} &(10) \end{aligned}$$

In order that $R_{tr}(\mathsf{SNR}) \doteq \mathsf{SNR} - O(\mathsf{SNR}^2)$, it requires that fraction of energy used in training to be as small as

$$\gamma \dot{\leq} \mathsf{SNR} \tag{11}$$

Combining (8) and (11), and recall that $E_{tr} = \gamma E_{total} = \gamma l \mathsf{SNR}$, we have

$$l \geq \mathsf{SNR}^{-3}$$

as a necessary condition that the near coherent throughput to be achieved.

3.2 Flash Training Schemes

Allowing flashiness in training schemes does not help to improve the performance at the coherent end ($\alpha = 1$). This is because to achieve the near coherent performance, it is necessary to use a $\delta \doteq \mathsf{SNR}^0$. To see that, suppose $\delta \doteq \mathsf{SNR}^\epsilon$ for $\epsilon > 0$, even in the AWGN channel, the resulting throughput is

$$\begin{aligned} \delta \log\left(1 + \frac{\mathsf{SNR}}{\delta}\right) &\leq \delta\left[\frac{\mathsf{SNR}}{\delta} - \frac{1}{2}\left(\frac{\mathsf{SNR}}{\delta}\right)^2 + \frac{1}{3}\left(\frac{\mathsf{SNR}}{\delta}\right)^3\right] \\ &= \mathsf{SNR} - O(\mathsf{SNR}^{2-\epsilon}) + O(\mathsf{SNR}^{3-2\delta}) \end{aligned}$$

for small ϵ, this is strictly less than $\mathsf{SNR} - O(\mathsf{SNR}^2)$. Intuitively, to achieve the coherent performance, one should transmit in all the available time frequency slots, thus the non-flashy training scheme is optimal, and $l \doteq \mathsf{SNR}^{-3}$ is still necessary.

However, at an intermediate level of coherence, $\alpha < 1$, allowing flashy training does reduce the gap between the performance of training schemes and the capacity. The following Lemma gives a characterization of this effect.

Lemma 2 *For a block fading channel with coherence time* $l \doteq \mathsf{SNR}^{-3\alpha}$*, using a flash training scheme with* $\delta \doteq \mathsf{SNR}^{-(1-\alpha)}$*, one can achieve a data rate of the order* $\mathsf{SNR} - O(\mathsf{SNR}^{1+\alpha})$

Proof:
Write the throughput of the described flashy training scheme as

$$I_{tr}(\mathsf{SNR}, \delta) = \delta I_{tr}\left(\frac{\mathsf{SNR}}{\delta}\right)$$

where $I_{tr}(\mathsf{SNR})$ is the throughput for a non-flashy training scheme with average power per symbol time as

$$\mathsf{SNR}' = \frac{\mathsf{SNR}}{\delta}$$

Now for $\delta \doteq \mathsf{SNR}^{-(1-\alpha)}$, we have

$$\mathsf{SNR}' \doteq \mathsf{SNR}^{\alpha}$$

Lemma 1 says that if the coherence time

$$l \doteq (\mathsf{SNR}')^{-3} = \mathsf{SNR}^{-3\alpha}$$

we have

$$I_{tr}(\mathsf{SNR}') \doteq \mathsf{SNR}' - O((\mathsf{SNR}')^2) = \mathsf{SNR}^{\alpha} - O(\mathsf{SNR}^{2\alpha})$$

and thus

$$\begin{aligned} I_{tr}(\mathsf{SNR}, \delta) &= \delta I_{tr}(\mathsf{SNR}') \\ &\doteq \mathsf{SNR} - O(\mathsf{SNR}^{1+\alpha}) \end{aligned}$$

is achievable. ○

Comparing to the non-flashy training schemes, which requires the coherence time to be of order $\mathsf{SNR}^{-(1+2\alpha)}$ to achieve the performance at coherence level α, the flash training schemes requires only $\mathsf{SNR}^{-3\alpha}$, which is significantly less for $\alpha < 1$. In particular, near the non-coherent end, the flash training scheme, with a slight variation that we randomly choose the blocks in which signals are transmitted, reduces to the on-off signaling schemes. Thus the capacity at $l = 1$ can also be achieved. The flash training schemes thus provide us a migration path between the two extremes.

It needs to be emphasized that the performance of the flash training schemes may not be optimal, in the sense that other schemes might require a shorter coherence time to achieve the same coherence level. Intuitively, the training schemes separates the channel estimation from the communication, thus a significant amount of energy is wasted. Joint channel estimation and communication schemes might thus perform significantly better when the channel is energy limited. The results on this issue will be reported in the journal version of this paper.

4 Conclusion

In this paper we studied the performance of training and flash training schemes, in fading channels with low SNR and long coherence time. We give a sufficient condition of how slow the channel can change over time in order that the near coherent performance can be achieved. We also demonstrate that the performance can be improved when training schemes are used in a flashy fashion. A notion of operational coherence level is also defined to bridge the gap between the coherent and the non-coherent extremes.

References

[1] R. S. Kennedy, *Fading Dispersive Communication Channels.* Wiley Interscience, New York, NY, 1969.

[2] I. E. Telatar and D. N. C. Tse, Capacity and mutual information of wideband multipath fading channels. *IEEE Trans. Info. Theory*, 46(4):1384–1400, July 2000.

[3] L.Zheng, and D.N.C.Tse, Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel. *IEEE Trans. Info. Theory*, 48(2):359–383, Feb 2002.

[4] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Info. Theory*, 48(6):1319 –1343, June 2002.

[5] M. Médard and R. Gallager, Bandwidth Scaling for Fading Multipath Channels *IEEE Trans. Info. Theory*, 48: 840-52, June 2002.

[6] Hassibi, B.; Hochwald, B.M. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Info. Theory*, 49: 951-63, April, 2003