6.962 Week 7 Summary:
Capacity of Fading Channels with Multiple Antennas

Presenter: Aaron Cohen

October 25, 2000

1 Introduction

We now turn our attention to the study of wireless systems. We will use a mobile phone system as the canonical example of such a system, which typically has many simultaneous users and is highly time-varying. This talk will focus on the single user case, with earlier and later talks shedding some light onto the difficult multi-user case. Under appropriate conditions, the time-varying nature can be modeled by a fading channel, where the received signal is simply a scaled version of the transmitted signal plus some additive noise.

In designing such a system, the most precious resources are bandwidth and power. We will show that using multiple antennas at both the transmitter and the receiver gives a dramatic increase in capacity for a fixed amount of bandwidth and power [1, 2, 3]. We will be primarily interested here in capacity results and not in efficient coding and equalization methods used to obtain rates close to capacity (see e.g. [4]).

We will first discuss fading channels in general. We will next look at the case of a single transmit and receive antenna. We will then examine the model proposed by Telatar [1] for the multi-antenna channel. We will finally look at the results obtained by Telatar. For more results see the very thorough review article [5]

2 Fading channels overview

This section is based on the presentation by Proakis [6, Chap. 14], and is intended to provide a brief introduction to the models and terminology widely used for wireless channels. We will assume throughout that we are using a modulation scheme with bandwidth W and signaling rate T_s where $W \approx 1/T_s$.

The time-varying nature of a wireless channel is represented by a random process $c(t;\tau)$ which gives the output of the channel at time t due to an impulse at time $\tau - t$ (plus some additive noise). Our first assumptions are that this random process is wide sense stationary (in t) and that responses at different values of τ are uncorrelated. Thus,

$$\frac{1}{2}E[c^*(\tau_1;t)c(\tau_2;t+\Delta t)] = \begin{cases} \phi_c(\tau_1;\Delta t) & \text{if } \tau_1 = \tau_2 \\ 0 & \text{otherwise} \end{cases},$$

and we call $\phi_c(\tau) \equiv \phi_c(\tau;0)$ the delay power spectrum of the channel. The range of values of τ where $\phi_c(\tau)$ is non-zero is called the multipath spread of the channel is denoted by T_m. Inverting this quantity gives us the *coherence bandwidth*,

$$(\Delta f)_c = \frac{1}{T_m},$$

which provides a crude measure of how different frequencies are affected by the channel. If $W \gg (\Delta f)_c$, then we say that the channel is *frequency-selective*. This usually results in severe distortion since different frequency bands are affected by different noise. Conversely, if $W \ll (\Delta f)_c$, then we say that the channel is *frequency-nonselective* or has *flat fading*. In this case, the entire signal is usually affected by a single multiplicative noise term.

We can also consider the correlation function of the Fourier transform $C(f;t)$ of $c(\tau;t)$ (taken with respect to τ with t fixed). In particular, let

$$\phi_C(\Delta f;\Delta t) = \frac{1}{2}E[C^*(f;t)C(f+\Delta f;t+\Delta t)]$$

and let $S_C(\lambda)$ be the Fourier transform of $\phi_C(0;\Delta t)$. Then, the Doppler spread B_d is the range of value of λ over which $S_C(\lambda)$ is non-zero. The inverse of the Doppler spread,

$$(\Delta t)_c = \frac{1}{B_d}$$

is the *coherence time* of the channel. This gives a crude measure of how long the impulse response remains constant. For example, if $T_s \ll (\Delta t)_c$, then the noise remains constant over a signaling interval.

If $(\Delta t)_c(\Delta f)_c \ll 1$, then we can design a system such that $W \ll (\Delta f)_c$ and $T_s \ll (\Delta t)_c$, which we call a slowly (flat) fading channel. We will focus exclusively on this case. In particular, we ignore the problem of inter-symbol interference (ISI).

3 Single antenna case

3.1 Model

In the single antenna case of a slowly fading channel, we can think of an equivalent baseband, discrete time system:

$$y_k = h_k x_k + n_k, \quad k = 1, 2, \ldots$$

where $\{x_k\}$ is the (complex) input sequence designed by the transmitter; $\{h_k\}$ is a (complex) random sequence which we will call the fading coefficients; and $\{n_k\}$ is a (complex) additive noise sequence. Typically, these three sequences are jointly independent.

The additive noise sequence is usually modeled as IID (complex) zero-mean Gaussian. Many different models for fading coefficients have been proposed. One of the most common is Rayleigh fading (other common distribution include Rician and Nakagami-m), which is appropriate when a large number of scatterers contribute to the fading. In this case, one can think of h_k with independent real and complex components which are both zero-mean Gaussian with identical variance. Equivalently, h_k has uniform phase with magnitude given by the Rayleigh pdf,

$$p_R(r) = \frac{2r}{\Omega} e^{-r^2/\Omega}, \; r \geq 0,$$

where $\Omega = E[R^2]$. The fading coefficients are typically correlated but h_k and h_{k+m} are approximately independent if $m \gg (\Delta t)_c / T_s$.

The power constraint is modeled by requiring that the input sequence $\{x_k\}$ satisfy

$$N^{-1} \sum_{k=1}^{N} |x_k|^2 \leq P$$

either in expectation (average power constraint) or with probability one (peak power constraint), where N is the blocklength. Any number of other constraints have also been examined.

It is sufficient to consider $E[|h_k|^2] = 1$ and $E[|n_k|^2] = 1$ and to vary P (the power constraint) in order to study all possible configurations with SNR $= P$.

3.2 Capacity results

We describe the capacity for three cases: fading coefficients known at receiver only, fading coefficients known at transmitter and receiver, and fading coefficients known at neither transmitter nor receiver. These scenarios are extreme points for many realistic situations where the transmitter and/or receiver might have some estimate of the fading coefficients (see e.g. [7]). Perfect knowledge of the fading coefficients is often called channel side information (CSI). In each of these case, we assume that $\{h_k\}$ form an IID sequence, but many of the same results apply if $\{h_k\}$ is stationary and ergodic with the same marginal distribution.

CSI at receiver only: For any distribution on the fading coefficients, the capacity achieving distribution on the inputs is Gaussian with power P. The resulting capacity[1] is given by [8]

$$C_{\mathrm{RCSI}}(1, 1, P) = E_h[\log(1 + |h|^2 P)], \tag{1}$$

where for consistent notation we will write $C(r, t, P)$ to denote the capacity with r receive antennas, t transmit antennas and SNR P. For Rayleigh fading, the capacity is given by [9]

$$C^{\mathrm{R}}_{\mathrm{RCSI}}(1, 1, P) = (\log e) e^{\frac{1}{P}} \int_{1/P}^{\infty} \frac{e^{-t}}{t}\, dt.$$

CSI at transmitter and receiver: The capacity achieving distribution is Gaussian with power given by the waterfilling algorithm [10]

$$P_x(h) = \begin{cases} P_0 - \frac{1}{|h|^2} & \text{if } |h|^2 \geq 1/P_0 \\ 0 & \text{otherwise} \end{cases},$$

where P_0 is chosen such that $E[P_x(h)] = P$. The resulting capacity (for every fading coefficient distribution) is

$$C_{\mathrm{TRCSI}}(1, 1, P) = E_h[\log(1 + |h|^2 P_x(h))].$$

[1] All capacities will be in units b/s/Hz and all logarithms will be base 2.

It is interesting to note that the difference between $C_{\text{RCSI}}(1,1,P)$ and $C_{\text{TRCSI}}(1,1,P)$ is quite small for Rayleigh and other typical fading distributions, particularly for large P [10].

No CSI: This case has been studied in [11] for Rayleigh fading. There, it is found that the capacity achieving distribution is discrete with a finite number of mass points (including one at zero). They find that there is little loss in capacity at low SNR (ratio between $C^{\text{R}}_{\text{RCSI}}(1,1,P)$ and $C^{\text{R}}_{\text{NoCSI}}(1,1,P)$ goes to one as P goes to zero), but that there is a great loss for high SNR.

Note that the previous discussion has focussed on the so called ergodic capacity, where we are free to choose the blocklength long enough to average over both the additive noise and the (possibly non-IID) multiplicative noise. Other notions suggested to measure performance for fading channels include capacity versus outage [9] and delay limited capacity [12].

4 Multi-antenna fading channel model

The paper by Telatar [1] considers the case where there are t transmit antennas and r receive antennas. At each time, the received sequence is given by

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n},$$

where $\boldsymbol{x} \in \mathbb{C}^t$ is the input to each of the transmit antennas, $\boldsymbol{y} \in \mathbb{C}^r$ is the output from each of the receive antennas, $\boldsymbol{n} \in \mathbb{C}^r$ is the additive noise at each of the receive antennas, and $H \in \mathbb{C}^{r \times t}$ is the fading matrix which describes the fading coefficient from each transmit antenna to each receive antenna. Both $\boldsymbol{n}$ and $\boldsymbol{H}$ are made up of IID complex (circular) Gaussian random variables with zero mean and variance 1. The power constraint is modeled by[2] $E[\boldsymbol{x}^\dagger \boldsymbol{x}] \leq P$, which is an average power constraint at each time. Note that this gives us a total power of P to allocate to the different transmit antennas, so that we can compare with the earlier results. Also note that the bandwidth used by both systems is the same. Some extra complexity might be needed at the transmitter and the receiver, but complexity is quite inexpensive these days. Telatar always assumes that $\boldsymbol{H}$ is known at the receiver (RCSI) and that an independent realization is drawn at each time step (the same results hold if stationary and ergodic).

[2]The superscript † means conjugate transpose.

4.1 Capacity results

He finds that the capacity achieving distribution is IID Gaussian over the t transmit antennas with each antenna using P/t of the power. The resulting capacity (for any IID fading matrix) is given by

$$C_{\mathrm{RCSI}}(r,t,P) = E_h[\log\det(I_r + (P/t)\boldsymbol{H}\boldsymbol{H}^\dagger)], \tag{2}$$

where I_r is the $r \times r$ identity matrix. Note the similarity between (1) and (2). When the fading is Rayleigh, he finds that the capacity is given by

$$C^{\mathrm{R}}_{\mathrm{RCSI}}(r,t,P) = \int_0^\infty \log(1+P\lambda/t) \sum_{k=0}^{\min\{r,t\}-1} \frac{k!}{(k+|r-t|)!} \left[L_k^{|r-t|}(\lambda)\right]^2 \lambda^{|r-t|} e^{-\lambda}\, d\lambda,$$

where L_j^i are the associated Laguerre polynomials. While not easily computed in general, one can easily compute this capacity for any given values of r, t, and P. we can verify that the formula given earlier for $C^{\mathrm{R}}_{\mathrm{RCSI}}(1,1,P)$ is the same as this one for $r=t=1$.

One significant property of this result is that for a fixed power P, the capacity scales linearly with the number of antennas, if both transmitter and receiver are using the same number of antennas. This is somewhat surprising since increasing the number of antennas at either the transmitter or receiver (but not both) results in no gain (asymptotically) or logarithmic gain, respectively. This property is a result of the distribution of the eigenvalues of the random matrix

$$\boldsymbol{W} = \begin{cases} \boldsymbol{H}\boldsymbol{H}^\dagger & \text{if } r < t \\ \boldsymbol{H}^\dagger\boldsymbol{H} & \text{if } t \geq r \end{cases}.$$

In particular, each of the $m = \min\{r,t\}$ eigenvalues $\boldsymbol{\lambda}_i$ of $\boldsymbol{W}$ has the same distribution and the distribution of $\boldsymbol{\lambda}_1/m$ converges (not to zero) as $m \to \infty$. Thus, one would suspect that the capacity would scale linearly with the number of transmitters for many distributions on $\boldsymbol{H}$, although the slope might be different for each one. These observations are based on rewriting (2) to get

$$C_{\mathrm{RCSI}}(r,t,P) = mE[\log\det(1 + (P/t)\boldsymbol{\lambda}_1)].$$

Another significant property is that at high SNR, each 3 dB increase in the SNR (i.e. doubling of the power) results in an increase in m b/s/Hz in capacity.

Telatar also describes the capacity when H is fixed. As in the single antenna TRCSI case above, the capacity achieving distribution is Gaussian using a waterfilling power distribution, but this time over the eigenvalues of W. One can extend this result to a multi-antenna channel with TRCSI, where now the random matrix $\boldsymbol{H}$ is known to both transmitter and receiver. In this case, the input distribution waterfills its power over both time and eigenvalues. However, as above, it has been suggested [13] that the gain from knowing $\boldsymbol{H}$ at the transmitter is not large, particularly at high SNR.

The final capacity result given by Telatar is the capacity versus outage when $\boldsymbol{H}$ is drawn according to the above distribution, and then fixed for all time. In this case, we can not guarantee that any positive rate is achievable with arbitrarily small probability of error. However, if we are allowed some fixed probability of outage, then we in fact can guarantee some positive rate. For some fixed probability P_{out}, this is computed by finding the largest rate such that the mutual information (viewed as a random variable in terms of the fading coefficients) is smaller than that rate with probability at most P_{out}. Telatar conjectures that if P_{out} is large enough, then we will not use all of the antennas to achieve the capacity versus outage.

4.2 Other models

We now give some variations on the model proposed by Telatar. The first is a model suggested by by Marzetta and Hochwald [2] for non-coherent communication. The second (proposed by some physicists from Bell Labs [14]) is a relaxation of Telatar's assumption that the coefficients in the fading matrix are IID.

4.2.1 Non-coherent model

In the papers by Marzetta and Hochwald [2] and Zheng and Tse [3], it is assumed that the receiver and the transmitter do not know that the fading coefficients, but that they stay constant for a coherence period of T signaling periods (i.e., $T \approx (\Delta t)_c/T_s$). During each coherence period, an independent realization of the fading coefficients is drawn. When $r = t = T = 1$, this is the problem addressed in [11] and discussed above (recall that the capacity achieving distribution was discrete in this case). For this model, we can write the output as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N},$$

where $\boldsymbol{H}$ is distributed as before, and where $\boldsymbol{Y} \in \mathbb{C}^{r \times T}$, $\boldsymbol{X} \in \mathbb{C}^{t \times T}$, and $\mathbb{C}^{r \times T}$ are matrices with columns representing what happens at each time and with rows representing what happens at each antenna. In [2], the capacity achieving distribution is found to be $\boldsymbol{X} \sim \boldsymbol{V\Phi}$, where $\boldsymbol{\Phi}$ is a $T \times T$ isotropically distributed unitary random matrix[3], $\boldsymbol{V}$ is a $t \times T$ real diagonal matrix, and $\boldsymbol{\Phi}$ and $\boldsymbol{V}$ are independent. As above, the optimal distribution on each diagonal element of $\boldsymbol{V}$ is discrete, but in most cases, it is sufficient to consider a deterministic matrix V with identical diagonal entries (known as unitary space-time modulation [15]). In [3], it is found that at high SNR each 3 dB increase in SNR gives an extra $m^*(1 - m^*/T)$ b/s/Hz in capacity, where $m^* = \min\{t, r, T/2\}$.

4.2.2 Non-IID fading coefficients

The assumption that the entries in the matrix $\boldsymbol{H}$ are IID is stated to be valid when all of the antennas have enough separation (on the order of a wavelength) within the transmit and receive arrays. However, a group of physicists at Bell Labs have looked at this problem and have proposed an alternative yet still simple model [14]. In particular, they suggest that each H_{ij} is still Rayleigh distributed, but that[4]

$$E\left[H_{i\alpha} H^*_{j\beta}\right] = A_{ij} B_{\alpha\beta},$$

for some positive definite Hermitian matrices $A \in \mathbb{C}^{t \times t}$ and $B \in \mathbb{C}^{r \times r}$. Note that this simplifies to the previous model when both A and B are identity matrices. One consequence of this result is that if $r = t$ and if the antennas are arranged in a cubic array with half wavelength spacing, then the capacity only increases as $r^{2/3}$ as opposed to linearly. For a linear or square array, the linear increase in capacity is still valid.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.

[2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[3] In such a matrix, each column is "uniformly" distributed over all vectors with norm 1, with the additional constraint that the columns are orthogonal.

[4] Jointly complex Gaussian random variables are uniquely specified by the covariance structure.

[3] L. Zheng and D. N. C. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the non-coherent mutli-antenna channel." Preprint, May 2000.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] E. Biglieri, J. Proakis, and S. Shamai, "Fading channes: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[6] J. G. Proakis, *Digital Communications.* McGraw-Hill, third ed., 1995.

[7] T. Klein, *Capacity of Gaussian Noise Channels with Side Information and Feedback.* PhD thesis, MIT, Camridge, MA, 2000.

[8] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inform. Theory*, vol. 16, pp. 353–355, May 1970.

[9] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular modile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[10] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[11] I. C. Abou Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inform. Theory*, 1998. submitted.

[12] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels – Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

[13] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, Autumn 1996.

[14] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. H. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 2000.

[15] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.