RESIDUAL NOISE AFTER INTERFERENCE CANCELLATION ON FADING MULTIPATH CHANNELS

R. G. Gallager

Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Dedicated to Tom Kailath, one of my oldest and best friends, on the celebration of his 60^{th} birthday. Much of the best early work on the fading multipath channels of interest in this paper was done by Tom [Ka59] while he was still a graduate student at MIT.

ABSTRACT

A popular information theoretic technique for multiaccess communication on a white Gaussian noise channel is to decode the users one by one. After each user is decoded, its encoded waveform is subtracted from the received signal, thus cancelling the interference from that user for the task of decoding subsequent users. This technique is not directly applicable to the fading multipath channels common to wireless communication. The problem is that what should be subtracted from the received signal is the response of the fading channel to the user's encoded waveform. Since the channel is unknown, the best that can be done is to subtract the convolution of the encoded waveform with an estimate of the channel. This leaves a residual noise term which is the convolution of the encoded waveform with the difference between the true channel and the estimated channel. The point of this paper is to show that this residual noise term is negligibly small for typical wireless situations.

1 INTRODUCTION

Cellular communication, personal communication systems (PCS), and packet radio systems all involve multiaccess communication, i.e., multiple transmitters sending data to the same receiver. Present day systems use a number of different techniques such as time division multiple access (TDMA), frequency division multiple access (FDMA), code division multiple access (CDMA), and collision resolution to control interference between the transmitters. The channels between these transmitters and receivers are typically fading multipath channels. For these channels, the response to each transmitted signal is a time-varying linear combination of delayed replicas of that transmitted signal, and the overall received signal is the sum of the responses to the various transmitters plus additive noise.

From an information theoretic standpoint, the above systems can be abstracted as multiaccess communication over time-varying multipath channels. In this abstraction, we restrict attention to a single receiver with multiple transmitters. The transmitters are limited to a common frequency band of width W and are each power constrained. The noise is modeled as white Gaussian, and the time-varying path strengths and delays are modeled as random processes. We can then attempt to find the multiaccess capacity region and achievable coding error probabilities as a function of rate and code block length.

For additive white Gaussian noise channels without time-varying multipath, the multiaccess capacity region is well known [CT91] and can be interpreted in terms of successive interference cancellation. That is, a set of m transmitter rates is within the capacity region if the receiver can decode the transmitters one by one. The first transmitter's code word is decoded with the other transmitters' code words treated as additional noise. The waveform for that code word is then subtracted from the received waveform, thus cancelling the interference of that waveform from successive decodings. The code words of subsequent transmitters are decoded and cancelled in the same way. The capacity region turns out to be the convex hull of the sets of rates decodable by this interference cancellation approach. Points in this convex hull can be achieved by time sharing between points achievable by interference cancellation. Time sharing has a number of system disadvantages for wireless systems [GJ91], but it turns out that arbitrary achievable points can also be achieved by interference cancellation directly. This can be done if some transmitters are conceptually split into two users, the available power of the transmitter and the required rate being split between the two users [RU96].

The situation for fading multipath channels is considerably more complex. As we shall see below, a fading multipath channel can be represented as an unknown time-varying linear filter. Thus the waveform from each of the m transmitters goes through an unknown time-varying linear filter. The received waveform is the sum of the outputs from these m filters along with white Gaussian noise. Suppose that we successfully decode a code word from one of the transmitters and attempt to do interference cancellation. We would like to subtract the convolution of the code word waveform and the time-varying linear filter from the received waveform, but since the time-varying filter is unknown, we instead use the convolution of the code word waveform and an estimate of the time-varying linear filter. Thus the interference cancellation is imperfect, and some residual noise, consisting of the code word waveform convolved with the difference between the true channel filter and the estimated channel filter, is left to interfere with subsequent decoding of other transmitters.

Our objective in this paper is to show that this residual noise is negligible for typical wireless situations. We will assume that each transmitter uses CDMA waveforms over a broad enough bandwidth that the interference from these transmitted signals can be modeled as white Gaussian noise over the bandwidth of interest. Because of this, we will study the residual noise in terms of a single transmitter, a single unknown filter, and additive Gaussian noise. We focus on the problem of estimating the unknown filter, and show that a Rake receiver [PG60] is an appropriate mechanism for both estimating the channel and detecting the transmitted signal. In the interest of simplicity, we shall make a number of simplifying assumptions as we proceed. Many of these assumptions can be avoided, but the results become less insightful.

Since we analyze only the issue of residual noise, we neglect many important problems associated with multiaccess fading multipath channels. One of these is the capacity region under the assumption that transmitters and receivers all know the fading multipath channels [CV93, Go94, KH95,TH96]. These analyses consider transmitters that dynamically change their power, spectral density, and/or rate as the channels change. Some results on the capacity when transmitters do not know the channels are contained in [Me95]. There is clearly a need for more work on the multiaccess capacity region of fading multipath channels under various feedback situations, but it seems clear that channel measurement must play a central role in this. We now proceed to analyze channel measurements and residual noise.

2 BASEBAND EQUIVALENTS

Consider M-ary signalling with the M signals $u_1(t), ..., u_M(t)$. Let T be the intersymbol duration, so that each T seconds, one of the signals $\{u_m(t), 1 \leq m \leq M\}$ is transmitted. We assume that $u_m(t)$ is essentially non-zero only for $-T \leq t \leq 0$ so that successive signals do not overlap. After passing through the multipath channel there will be some overlap which we discuss later. The signals all have bandwidth W, centered around some carrier frequency $f_0 \gg W$. Let $U_m(f) = \int u_m(t)e^{-j2\pi ft}dt$ be the Fourier transform of $u_m(t)$ for each m, and define the baseband equivalent waveforms $x_m(t)$ in terms of their Fourier transforms $X_m(f)$ where $X_m(f) = U_m(f + f_0)$ for $f > -f_0$ and $X_m(f) = 0$ otherwise (see Fig. 3.1).

This way of going from passband to baseband is not entirely conventional. In particular,

$$u_m(t) = 2\mathrm{Re}[x_m(t)e^{j2\pi f_0 t}] = 2\mathrm{Re}[x_m(t)]\cos(2\pi f_0 t) - 2\mathrm{Im}[x_m(t)]\sin(2\pi f_0 t) \tag{3.1}$$

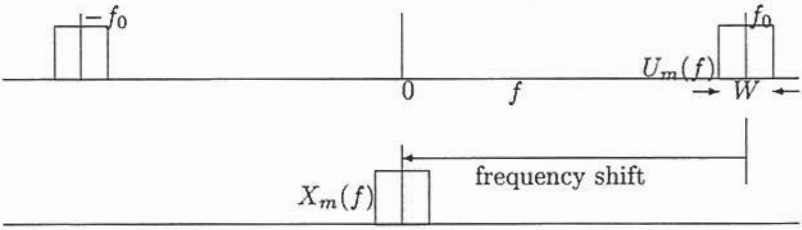

Figure 3.1 Baseband equivalent; $U_m(f)$ and $X_m(f)$ are generally complex.

Also, if $\mathcal{E}$ is the energy in the signal $u_m(t)$, then

$$\mathcal{E} = \int_{-\infty}^{\infty} [u_m(t)]^2 \, dt = 2 \int_{-\infty}^{\infty} |x_m(t)|^2 dt \,. \tag{3.2}$$

The desirable feature of this scaling is that if $u(t)$ is passed through a linear filter of impulse response $g(t)$ to get output $v(t)$, and if $x(t)$, $h(t)$, and $y(t)$ are the corresponding baseband equivalents of $u(t)$, $g(t)$, and $v(t)$, then

$$v(t) = u(t) * g(t) \quad ; \quad V(f) = G(f)V(f)$$

$$y(t) = h(t) * x(t) \quad ; \quad Y(f) = H(f)X(f) \tag{3.3}$$

Since the arguments to follow depend critically on being able to view signals and filters interchangeably, we have defined baseband waveforms so that (3.3) is satisfied, and by necessity this forces the peculiar energy scaling in (3.2).

Assume that the real and imaginary parts of $x_m(t); 1 \le m \le M$ are pseudo noise signals (as used in CDMA). These signals have the property that $|U_m(f)|$ is essentially constant over the signalling bandwidth $|f - f_0| \le W/2$. It follows that $|X_m(f)|$ is essentially constant over $|f| \le W/2$ and 0 elsewhere. For simplicity, we henceforth assume that $|X_m(f)|$ is exactly constant over $|f| \le W/2$ and zero elsewhere. Applying Parseval's equation to (3.2), $\int_{-\infty}^{\infty} |X_m(f)|^2 df = \mathcal{E}/2$, so

$$|X_m(f)|^2 = \frac{\mathcal{E}}{2W} \quad \text{for} |f| \le W/2 \quad ; \quad |X_m(f)|^2 = 0 \quad \text{for} |f| > W/2 \tag{3.4}$$

Since $|X_m(f)|^2$ and $R_m(t) = \int x_m^*(\tau)x_m(t+\tau)d\tau$ are Fourier transforms, it follows from (3.4) that $R_m(t) = (\mathcal{E}/2)\sin(\pi W t)/(\pi W t)$. Thus, if we view $x_m(t)$ in terms of its samples $x_{m,i}$ at rate W, we have $\sum_i x_{m,i}^* x_{m,i+j} = (\mathcal{E}W/2)\delta(j)$.

It is not possible to find waveforms $x(t)$ that are both time limited to the signal interval T and low pass limited to the band $W/2$. CDMA systems, however, have a relatively large time bandwidth product, $WT \gg 1$ (which is why they are called spread spectrum systems), and for this reason, waveforms can be found that are both approximately time limited and frequency limited. Finding such waveforms with desirable cross correlation properties is a large and very well studied field, but studying this would draw us away from our main purpose. Thus in what follows, we simply assume (3.4) to be valid, and recognize that the approximation can be quite good for $WT \gg 1$.

3 THE EFFECT OF MULTIPATH

Let $\tau_i(t)$ be the propagation delay of the i^{th} propagation path at time t, and let $a_i(t)$ be the strength of that path, at least within the frequency range of interest around f_0. Both a_i and τ_i change slowly with time. The impulse response of the channel, i.e., the output at time t due to an impulse τ seconds earlier is then $g(\tau,t) = \sum_i a_i(t)\delta(\tau - \tau_i(t))$. Thus the response to a signal u(t) is

$$v(t) = \int u(t-\tau)g(\tau,t)d\tau = \sum_i u(t-\tau_i(t))a_i(t) \qquad (3.5)$$

Defining $G(f,t) = \int g(\tau,t)e^{-j2\pi f\tau}d\tau$, we have $G(f,t) = \sum_i a_i(t)e^{-j2\pi f\tau_i(t)}$, again within the frequency range of interest. Define $H(f,t) = G(f+f_0,t)$ for the baseband region of interest. Then

$$H(f,t) = \sum_i a_i(t)e^{-j2\pi f_0\tau_i(t)}e^{-j2\pi f\tau_i(t)} = \sum_i \alpha_i(t)e^{-j2\pi f\tau_i(t)} \qquad (3.6)$$

$$\text{where} \quad \alpha_i(t) = a_i(t)e^{-j2\pi f_0\tau_i(t)}$$

Letting $x(t)$ and $y(t)$ be the baseband equivalents of $u(t)$ and $v(t)$ respectively, and letting $X(f)$ and $Y(f)$ be the corresponding Fourier transforms, it can be shown after a little manipulation that

$$y(t) = \int_{-\infty}^{\infty} X(f)H(f,t)e^{j2\pi ft}df \qquad (3.7)$$

This shows that $H(f,t)$ for $|f| > W/2$ has no effect on the output. Thus we arbitrarily define $H(f,t)$ to be 0 for $|f| > W/2$. Inverse Fourier transforming (3.6) with this modification, the baseband equivalent filter is

$$h(\tau,t) = \sum_i \alpha_i(t)\frac{\sin[\pi W(t-\tau_i(t))]}{\pi(t-\tau_i(t))} \qquad (3.8)$$

Also, transforming the right hand side of (3.7), we get

$$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(\tau,t)d\tau \qquad (3.9)$$

Note that $h(\tau, t)$ has one filtered impulse for each path, and that the sinc pulse representing the filtered impulse has a peak that increases with W and a width that decreases, thus keeping unit area. The multipath structure does not change as the bandwidth of the input is changed, but the filter $h(\tau, t)$ does change, since $h(\tau, t)$ represents only the effect of the channel over the given bandwidth. This is an important point, since the effect of the channel is typically very complex, but we need measure its effect only on the signals in the given band. Since we want to measure the channel over the bandwidth $W/2$, we want to characterize it in the simplest way over that band.

Define L as the multipath spread of the channel; this is the difference in propagation delay between the longest and shortest path. For W large, $h(\tau, t)$ is approximately 0 for $\tau < 0$ and for $\tau > L$ (in communication, one usually adjusts the time reference at the receiver to be delayed from that at the transmitter by the shortest (or sometimes the strongest) propagation delay). For smaller W, it can be seen that $h(\tau, t)$ is non-zero over an interval L' consisting of L plus several times $1/W$. For cellular mobile communication, L is typically between a few microseconds and 30 microseconds, and for PCS, L is typically much smaller, on the order of 100 nsec. If $L = 10$ μsec, and $W = 10^6$ H, then $h(\tau, t)$ could be represented (through the sampling theorem) by slightly more than 10 samples in τ; each sample is complex, so measuring h at any given time corresponds to measuring slightly more than 20 real numbers.

Define B as the Doppler spread of the channel; this is the difference between the largest and the smallest Doppler shift. Typical values in a mobile system are around 100 H. B determines how quickly $h(\tau, t)$ can change with t. The phase in the path strength $\alpha_i(t)$ can change significantly over the interval $1/B$, so that measurements of the channel become outdated over intervals of duration $1/B$. We will assume in what follows that the signalling interval T is very much smaller than $1/B$, and thus we assume that $h(\tau, t)$ is constant as a function of t over a signal interval T. Thus $h(\tau, t)$ is modeled as a linear time invariant filter over individual signal intervals, allowing one to play all the games of elementary linear systems. One must recognize, of course, that $h(\tau, t)$ changes significantly over many signalling intervals, so that one cannot simply measure h once and for all.

4 ESTIMATING $H(\tau, T)$

First ignore noise, assume that $x_m(\tau)$ is transmitted, and consider passing the channel output, $x_m(\tau) * h(\tau, t)$ through a filter matched to x_m (i.e., a filter with impulse response $x_m^*(-\tau)$) (see Fig. 3.2).

Taking Fourier transforms, we have

$$R_m(f) = X_m(f)H(f,t)X_m^*(f) = |X_m(f)|^2 H(f,t) = \frac{\mathcal{E}}{2W} H(f,t) \qquad (3.10)$$

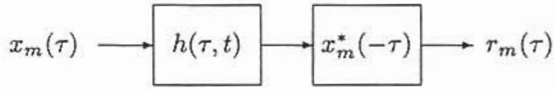

Figure 3.2

where we have used (3.4). Taking the inverse Fourier transform, we see that $r_m(\tau) = \frac{\mathcal{E}}{2W} h(\tau, t)$. Since we are looking at an input in the interval $[-T, 0]$, and we are assuming that $h(\tau, t)$ does not change in t over intervals of duration T, the parameter t can be taken to be 0. This suggests that the output should be attenuated by $2W/\mathcal{E}$ in order to obtain an estimate of $h(\tau, t)$ at t=0.

We now put the white noise back in the picture and look at the output of the attenuated matched filter including noise (see Fig. 3.3). Assume the noise has spectral density $N_0/2$. Filtering the noise to $|f - f_0| \leq W/2$, and defining the baseband equivalent noise, as the upper sideband shifted down by f_0, the baseband equivalent noise process is complex Gaussian and has the spectral density $N_0/2$ for $|f| \leq W/2$. It follows that the noise power of the baseband waveform is $N_0 W/2$, which is half the noise power of the band pass waveform. Thus we have scaled the noise in the same way as the signal. Physically, when one demodulates a passband waveform into quadrature baseband components, one can scale those baseband waveforms arbitrarily, but the signal and noise must be scaled the same way.

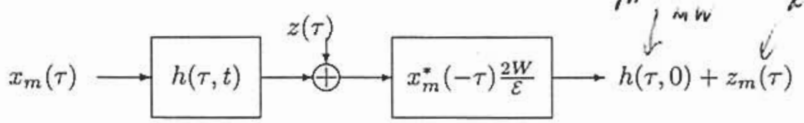

Figure 3.3

We have seen that the component of the output due to the signal $x_m(\tau)$ is $h(\tau, 0)$. To analyze a sample $z_m(\tau)$ of the output noise process, note that the Fourier transform of the attenuated matched filter is $X_m^*(f) 2W/\mathcal{E}$. The spectral density of the filter, i.e., the magnitude squared of the Fourier transform, is $\mathcal{E}/(2W))[2W/\mathcal{E}]^2 = 2W/\mathcal{E}$. Since the input process has the spectral density $N_0/2$, the output $\{z_m(\tau)\}$ comes from a process with the spectral density $N_0 W/\mathcal{E}$. Since this output noise process has bandwidth $W/2$, $z_m(\tau)$, for any given τ is a sample value of a random variable of variance $N_0 W^2/\mathcal{E}$, so

$$\text{Var}[z_m(t)] = N_0 W^2/\mathcal{E} \tag{3.11}$$

Now consider a rake receiver (see Fig. 3.4). If $h(\tau, t)$ is known, then the optimal detector for the M-ary signal set $x_m(\tau)$, $1 \leq m \leq M$, through the filter $h(\tau, t)$, is simply a set of matched filters matched to the convolution of x_m and h, i.e., $x_m^*(-\tau) * h^*(-\tau, t)$. The decision on m in Fig. 3.4 replaces $h^*(-\tau, t)$ with $\hat{h}^*(-\tau, t)$, which is reasonable if the estimate is good. Assume that $h(\tau, t)$ is well estimated, and that a correct decision is made on the input m. Given this decision, the output of the filter matched to the signal x_m yields a new estimate of h plus additive Gaussian noise. The device to estimate h then uses the decision on m to accept the output from the m^{th} matched filter to update the old estimate of h.

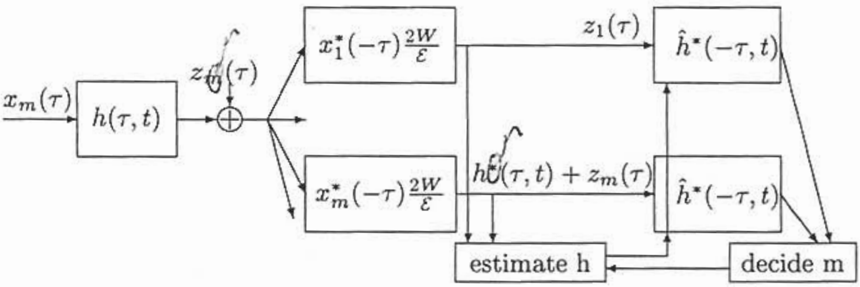

Figure 3.4 Rake receiver.

To avoid worrying about the optimal estimate of h, we can get an approximate answer by assuming that the estimate $\hat{h}$ of h is simply the linear average of the previous n measurements. Here one measurement is made each T seconds and the i^{th} such measurement, made at time $-iT$, comes from matched filter m_i, where signal m_i was sent at that time. Since the channel filter remains almost constant for a time on the order of $1/B$, we take $n = 1/(BT)$. Let $z'(\tau)$ be the error in $\hat{h}(\tau, 0)$, i.e., $\hat{h}(\tau, t) = h(\tau, t) + z'(\tau)$. Since taking an average over n measurements with IID noise reduces the noise variance by a factor of n, it can be shown that

$$\text{Var}(z'(\tau)) = \frac{N_0 W^2}{n\mathcal{E}} = \frac{BTN_0W^2}{\mathcal{E}} \quad ; \quad 0 \leq \tau \leq L' \tag{3.12}$$

We assume that the multipath spread L' (including the limitation to bandwidth $W/2$) is known, and thus that $\hat{h}(\tau, t)$ is taken to be 0 for $\tau < 0$ and $\tau > L'$. Since all of the noise processes being averaged are white over $|f| \leq W/2$, $\{z'(\tau)\}$ is a sample function of a process that is white over $|f| \leq W/2$ and is non-zero only over the interval $[0, L']$.

5 RESIDUAL NOISE

Finally we have the problem of determining the residual noise if the effect of the detected signal is subtracted from the received waveform (again assuming the correct signal was detected). The effect of the signal $x_m(\tau)$ on the received signal is $x_m(\tau) * h(\tau, t)$. The quantity subtracted from the received signal in cancelling the interference from this user is $x_m(\tau) * \hat{h}(\tau, t)$, and thus the residual noise after interference cancellation is $\phi(\tau) = x_m(\tau) * z'(\tau)$. Taking Fourier transforms, $\Phi(f) = X_m(f)Z'(f)$ and

$$|\Phi(f)|^2 = |X_m(f)|^2|Z'(f)|^2 = \frac{\mathcal{E}}{2W}|Z'(f)|^2 \tag{3.13}$$

Thus,

$$\int \frac{\mathcal{E}}{2W}|Z'(f)|^2 df = \int |\phi(\tau)|^2 d\tau = \int \frac{\mathcal{E}}{2W}|z'(\tau)|^2 d\tau \tag{3.14}$$

Taking the expected value of the final terms in (3.14),

$$\int \mathrm{Var}|\phi(\tau)|d\tau = \int \frac{\mathcal{E}}{2W}\mathrm{Var}|z'(\tau)|d\tau = \frac{BTN_0WL'}{2} \tag{3.15}$$

This is the baseband expected energy of the residual noise in the band $|f| \leq W/2$ and over the interval $0 \leq t \leq T$. Since $z'(\tau)$ is white over the band $|f| \leq W/2$, $\phi(\tau)$ is also white over $|f| \leq W/2$. Thus the spectral density of this noise power (averaged over the time interval $(0, T)$) is $BL'N_0/2$. Since the spreading product BL' is small for most wireless situations, this indicates that the residual noise is small relative to the ordinary additive noise of spectral density $N_0/2$. When multiaccess communication is taken into account, the noise that effects the filter measurement becomes not only the white noise but also the other users signals, which have been passed through their own multipath filters before contributing to the measurement of the filter in question.

6 DISCUSSION

The analysis here indicates that the residual noise is a factor BL' of the background noise and the interference from other users, and rather surprisingly, it does not depend on either the signal power or the signal bandwidth. The assumption, however, was that the signals could be successfully detected, and, of course, successful detection does depend on signal power and bandwidth. In fact, it can be shown that pseudo-noise signals of any given power and duration cannot be spread arbitrarily in bandwidth and still be detected on a fading multipath channel. This effect appears only indirectly here, where in it can be seen that the estimate of the channel becomes poor as W becomes large. This increasingly poor estimate does not increase the residual noise spectral density because the signal spectral density is decreasing with W, but it does increase the difficulty of detection.

It is not difficult to actually analyze the estimation error in estimating h. For any fixed τ, if we know the correlation function in t for the random process with sample functions $h(\tau, t)$, then we can use discrete Kalman filtering to find the minimum mean square error linear estimate. One finds, on doing this, that our assumption of $(BT)^{-1}$ estimates with IID noise is very optimistic, but it should be clear that the result only changes by a scale factor. A more serious issue is that with coding, there is significant delay before symbols can be correctly decoded, and this increases the delay in estimating the channel. One can imagine an iterative approach where symbols are detected without delay, this is used to update the channel estimate, and then a better channel estimate is made after decoding. One could also use a Viterbi decoder where different channel estimates are carried along with different potential paths, but this is not very attractive.

Perhaps the best way to look at this is that whatever method is used in decoding a user, that decoded data can be used to estimate the channel and cancel interference if decoding is correct. This can introduce large delays overall, because each user is delayed until the interference cancellation is done for the earlier users to be decoded, but interference cancellation is possible whenever decoding is possible.

It appears to be possible to carry through the analysis without assuming a flat spectral density for the input waveforms, but this seems to be an exercise for people who like complex calculations. Finally, we have ignored inter-symbol interference for the transmitter being detected. This seems to be reasonable if L' is small relative to the signal interval T. It is also reasonable if the spreading factor WT is large, since then the neighboring symbols appear like noise spread over the band, and act much like the interference from other users.

In summary, we must ask whether interference cancellation might someday be practical for wireless communication. We have shown that residual noise is not a major problem there, but delay might be an insurmountable problem, and interence from users in other cells might be sufficiently large that it doesn't pay to cancel interference within a cell. Thus the question is still open.

Acknowledgements

This research was supported in part by ARO under grant DAAH04-95-1-0103, in part by NSF under grant 9314341-NCR, and in part by ARO under grant DAAL-03-92-G-115. A rough draft of the basic results appeared in July 1994 as a LIDS report, LIDS-P-2254, M.I.T.

REFERENCES

[CV93] R.S.Cheng, S.Verdu, "Gaussian Multiaccess Channels with Capacity Region and Multiuser Water-filling," IEEE Trans. IT, Vol 39, No. 3, May 1993.

[CT91] T.Cover and J.Thomas, *Elements of Information Theory*, Wiley, 1991.

[GJ91] K.S. Gilhausen, I.M.Jacobs, R.Padovani, A.J.Viterbi, L.A.Weaver, and C.E.Wheatley, "On the Capacity of a cellular CDMA system," IEEE Trans. Vehic. Tech., Vol 40, No.2, May 1991.

[Go94] A.Goldsmith, "Design and Performance of High-speed Communication Systems over time-varying radio channels," Ph.D. thesis, University of California, Berkeley, 1994.

[Ka59] T. Kailath, "Sampling Models for Linear Time-Variant Filters," TR 352, MIT Research Laboratory of Electronics, May 25, 1959.

[KH95] R.Knopp, P.H.Humblet, "Channel Control and Multiple-Access, ICC, Seattle, Wash.., June 1995

[Me95] M.Medard, "The Capacity of Time-Varying Multiple User Channels in Wireless Communication," Ph.D. thesis, MIT, 1996.

[PG60] R.Price and P.E.Green, "Signal Processing in Radar Astronomy - Communication via Fluctuating Multipath Media," TR 234, Lincoln Laboratory, MIT, 6 Oct. 1960.

[RU96] B.Rimoldi and R.Urbanke, "A Rate Splitting approach to the Gaussian Multiple-Access Channel," IEEE Trans. IT Vol 42, No. 2, March 1996.

[TH96] D.Tse, S.Hanly, "Capacity Region of the Multi-access Fading Channel under Dynamic resource allocation and Polymatroid optimization," IEEE Information Theory Workshop, Haifa Israel, June 1996.