An Adaptive Antenna Integrated with Automatic
Gain Control for a Receiver Front End

by

Joel L. Dawson

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

May 21, 1997

Copyright 1997 Joel L. Dawson. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and
distribute publicly paper and electronic copies of this thesis
and to grant others the right to do so.

Author

Department of Electrical Engineering and Computer Science
May 21, 1997

Certified by

Professor James K. Roberge
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Department Committee on Graduate Theses
An Adaptive Antenna Integrated with Automatic
Gain Control for a Receiver Front End

by

Joel L. Dawson

Submitted to the
Department of Electrical Engineering and Computer Science

May 21, 1997

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The concept of antenna diversity has been widely employed to improve microwave receiver performance in a dynamic RF environment. Here, a new type of "smart" antenna has been developed, which utilizes two antenna elements. The final design represents a novel, integrated solution to two of the problems faced by mobile receivers as a result of multipath propagation: (1) information loss due to fading; (2) front end saturation due to unexpected peaks in received signal intensity. The nonlinear, analog, adaptive feedback controller at the heart of the system is described and analyzed, and the results of testing with the final RF prototype are presented.

Thesis Supervisor: James K. Roberge
Title: Professor of Electrical Engineering
To my mother

She openeth her mouth with wisdom; and in her tongue is the law of kindness.

and to my father

...his delight is in the law of the Lord; and in his law doth he meditate day and night.
# Table of Contents

List of Figures .................................................................................................................. 5

List of Tables ..................................................................................................................... 6

Chapter 1: Introduction to the Problem ............................................................................. 7
  The Structure of a Modern Microwave Receiver ........................................................... 8
  The RF Environment ......................................................................................................... 12
  The Impact of the RF Environment on Receiver Design .............................................. 18
  The Proposed System ....................................................................................................... 22
Chapter 1 References ......................................................................................................... 24

Chapter 2: The Theory Behind the Method ....................................................................... 26
  Methods of Spatial Combining .......................................................................................... 26
  The Problem of Designing a Controller ........................................................................... 34
Chapter 2 References ......................................................................................................... 38

Chapter 3: Design and Analysis of the Controller ............................................................. 39
  Overview of the Circuit ...................................................................................................... 39
  Design of the Non-Adaptive Circuit ............................................................................... 41
  Testing and Analysis of Non-Adaptive Circuit ............................................................... 47
  Adaptive Behavior, Part I: Fast Transient Response ..................................................... 56
  Adaptive Behavior, Part II: Adaptation on a Longer Time Scale .................................... 60
  Analog Processing vs. Digital Processing ......................................................................... 64

Chapter 4: Completing the RF Prototype .......................................................................... 66
  Overview of the RF system; Design of the Phase Shifter ............................................... 66
  Design of the Power Detector ........................................................................................... 72
Chapter 4 References ......................................................................................................... 74

Chapter 5: Experiment, and Results .................................................................................. 75
  The RF Experiment .......................................................................................................... 76
  Results and Discussion ..................................................................................................... 78

Chapter 6: Conclusion .......................................................................................................... 81

Acknowledgments .............................................................................................................. 85
List of Figures

Figure 1-1: Constituent blocks of a modern microwave receiver 8
Figure 1-2: A common LNA topology 9
Figure 1-3: The impulse response of an imperfect communications channel 19
Figure 1-4a: An equalizer 19
Figure 1-4b: The effect of a well-designed equalizer on the received signal 19
Figure 1-5: An electronically variable attenuator 21
Figure 1-6: The proposed system 22
Figure 2-1: Probability that a fade of depth x will occur for various values of M 29
Figure 2-2: A maximal ratio combining system 31
Figure 2-3: A comparison of three combining methods 32
Figure 2-4: Block diagram of controller 36
Figure 3-1: Your basic controller 40
Figure 3-2: Comparison of command and disturbance transfer functions 41
Figure 3-3: Detailed schematic of non-adaptive circuit 42
Figure 3-4: Controller output for a DC input 49
Figure 3-5: Controller output for a DC input (smaller time scale) 50
Figure 3-6: Step response of controller 51
Figure 3-7: Response to a ramp input 51
Figures 3-8 through 3-11: Controller tracking behavior 52
Figure 3-12: The signal at system ground 53
Figure 3-13: Greatly exceeding the triangle wave cutoff frequency 53
Figures 3-14 through 3-18: Sinusoidal response 54-55
Figure 3-19: I.boost current source 57
Figure 3-20: Fast transient detect 58
Figures 3-21 through 3-22: Improved step recovery time 59
Figures 3-23 through 3-25: New sinusoidal behavior 59-60
Figure 3-26: Slow adaptation circuit 61
Figures 3-27 through 3-31: Improved sinusoidal response 63
Figure 3-32: Block diagram of final controller 64
Figure 4-1: Placement of phase shifter and power detector 67
Figure 4-2: Simplified diagram of phase shifter 68
Figure 4-3: Detailed schematic of phase shifter 70
Figure 4-4: Simplified diagram of power detector 72
Figure 4-5: Power detector used in prototype 73
Figure 5-1: RF experiment 76
Figure 5-2: Power command set too high 79
Figure 5-3: Power command achievable for a fraction of each modulation cycle 79
Figures 5-4 through 5-7: Summary of system performance in various situations 80
List of Tables

Table 1-1: Time scales for fading events 17
Table 5-1: Results of the fading experiment 78
Chapter 1: Introduction to the Problem

In recent years, there has been a surge in the demand for wireless communications devices. Producers of cellular phones and personal pagers enjoy a burgeoning market for their products; in some ways, then, this surge is indicative of a growing consumer demand for convenience. Fundamentally, however, it reflects the expectation that wireless technology, once perfected to the point of providing robust communications at a reasonable cost, will enable its users to save money. It is clear that the leading provider of such communications also stands to reap considerable economic rewards. This realization has sparked a great deal of commercial interest, and a corresponding great deal of research interest, in the field of wireless technology.

Consider the problem of establishing a local area network (LAN). Software, hardware, and cabling all contribute to the cost of this investment. Cabling can account for as much as 40% of the total [1], and is the least flexible part of the entire setup. Hardware can be moved, as can software, but cabling is fixed in its location. Moreover, the types of cabling installed can limit the extent to which a network can be reconfigured. The cost of moving a cabled LAN sometimes approaches that of a new installation. All told, the cost to the U.S. industry of relocating LAN terminals has been estimated to be in the billions of dollars.
Microwave engineers, faced with the task of developing superior transceiver technology, are beset by an unforgiving set of constraints: the devices must be inexpensive, yet operate at high frequencies (tens of gigahertz) and work well in an extremely dynamic RF environment. The proposed system is intended to improve a transceiver's ability to receive signals. In the following sections and chapters, therefore, the discussion of modern microwave technology will be focused on receiver design.

**The Structure of a Modern Microwave Receiver**

A block diagram of a typical receiver is shown in figure 1-1.

![Block Diagram of a Modern Microwave Receiver](image)

**Figure 1-1: Constituent blocks of a modern microwave receiver**

A full appreciation of the problem at hand requires a general knowledge of the structure of a modern microwave receiver. Each of the pictured system blocks is discussed in turn in the following paragraphs.

The low-noise amplifier, or LNA, is typically the first active device in the signal path of a microwave receiver. It has two principal figures of merit: power gain and noise figure. The importance of power gain represents a significant departure from low-frequency amplifier design, where voltages and currents tend to be the variables of interest. An ideal multistage voltage amplifier, for instance, is composed of gain blocks with infinite input impedance and zero output impedance. The output current from each stage is zero, as is the power transfer from
stage to stage. But an electromagnetic wave propagating through space cannot be fundamentally tied to a voltage or a current; energy is traveling in the form of electric and magnetic fields, and the only meaningful characterization of signal strength is the associated Poynting flux. Accordingly, it is the task of the RF front end in general, and of the LNA in particular, to magnify signal power while adding as little noise as possible. The topology for a typical LNA is shown in figure 1-2.

![Figure 1-2: A common LNA topology](image)

In low-frequency systems, where voltages (or perhaps currents) are the variables of interest, certain constraints exist on the input and output impedances of the system blocks. The same can accurately be said of microwave circuits, though the nature of the constraints is different. In the case of the former, it is desirable to have an extreme impedance mismatch between the output of one block and the input of the next to minimize "loading." In the case of the latter, inputs are conjugate matched to outputs to maximize the power transfer. The matching networks in figure 1-2 are usually comprised of passive, purely reactive elements, and are placed in the circuit as impedance transformers. Thus, they serve a role analogous to voltage and current buffers in low-frequency designs.

The gain element in figure 1-2 is typically a single transistor. A variety of bipolar and field-effect transistors have been developed for RF purposes, with GaAs devices favored in extremely high frequency applications. Field-effect devices tend to be faster and less noisy than
their bipolar counterparts. However, they are often harder to match and to stabilize: at times, loss must be purposely introduced into the matching network to prevent oscillation. This loss, particularly if inserted in the input matching network, can result in noise performance that is far worse than the theoretical limit specified on the device data sheet. The careful designer will hesitate, therefore, before choosing a FET over a BJT solely for noise reasons.

It is crucial to the performance of the RF front end that the first amplification stage degrade the signal-to-noise ratio (SNR) as little as possible. Qualitatively, the noise figure is a measure of this degradation. Quantitatively, it is defined as [3]

\[
F = \frac{S_i}{N_i} \cdot \frac{S_o}{N_o}
\]  

(1-1)

where S and N denote signal and noise signal powers, respectively, and the subscripts \(i\) and \(o\) refer to input and output. The input noise power is taken to be that of a matched resistor at 290 K. The noise characteristics of any cascade of stages are dominated by those of the first stage; this is why the design of the low-noise amplifier is so critical. It can be shown that the overall noise figure for a cascade of amplifiers is given by

\[
F_{\text{cas}} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots
\]  

(1-2)

where \(G_n\) is the gain of the \(n\)th stage. It is clear from (1-2) that a high-gain, low-noise first stage goes a long way towards ensuring good noise performance for the entire front end. The LNA effectively determines the ability of the receiver to pick out weak signals from a noisy environment, a figure of merit referred to as the sensitivity.

The development of good bandpass filters is, itself, an entire field of research. We will not dwell on them here, except to note that in general microwave systems are narrowband. This is a consequence of a crowded RF spectrum; specific wireless applications are allocated narrow bands by the FCC, and to venture outside that band is to invite interference problems (and, as it happens, legal problems) of considerable proportions. There remains a great demand for bandpass filters with solid out-of-band rejection and low insertion loss.
Components designed to operate at radio frequencies are expensive and difficult to work with. Once the signal has been acquired, therefore, it is desirable to perform a “down conversion” of the carrier to a more manageable intermediate frequency; this is the task of the mixer. It is well known that the multiplication of two pure sinusoids yields two resultant sinusoids: one at the difference of the two frequencies, and one at the sum. Accordingly, one strategy is to multiply the incoming signal by a locally generated sinusoid that is close in frequency to the carrier. By filtering out the sum component, we are left with a much more manageable IF carrier. In many actual systems a true multiplication is not used. Rather, the received signal and that of the local oscillator are added together and then passed through a nonlinearity. This is best shown mathematically.

\[ V_{\text{sig}} = e^{j\omega_1 t} + e^{-j\omega_1 t} \]  
\[ V_{\text{LO}} = e^{j\omega_2 t} + e^{-j\omega_2 t} \]  
\[ V_{\text{in}} = V_{\text{sig}} + V_{\text{LO}} \]  
\[ V_{\text{out}} = a_0 + a_1 V_{\text{in}} + a_2 V_{\text{in}}^2 + a_3 V_{\text{in}}^3 + \cdots \]

Zero order term: \( a_0 \)

First order term: \( 2a_1(\cos \omega_1 t + \cos \omega_2 t) \)

Second order term: \( a_2 \left[ 4 + 2(\cos 2\omega_1 t + \cos 2\omega_2 t) + 4(\cos(\omega_1 - \omega_2) t + \cos(\omega_1 + \omega_2) t) \right] \)

We see that the quadratic term in the expansion of the transfer characteristic yields the desired sum and difference terms. It also results in harmonics (at frequencies \( 2\omega_1 \) and \( 2\omega_2 \)), but given that they are close in frequency to the sum term, it stands to reason that they can be easily filtered out. In dealing with the third order term, however, we find that we are not so fortunate. Among the several frequency components added by this term are sinusoids at \( 2\omega_1 - \omega_2 \) and \( 2\omega_2 - \omega_1 \) which are notoriously difficult to remove. A mixer designer is thus burdened with constraints on the nonlinear terms of the transfer characteristic, and we see now why the design of mixers can be challenging. It turns out that the cubic nature of the third order term works in the designer’s favor in the case of small amplitude signals, for which the lower order terms dominate. Still, the
engineer must bear in mind that there is a maximum input level beyond which the aforementioned intermodulation products, as they are called, render the mixer useless.

It is worth noting that a nonlinearity anywhere in the system has the potential to behave this way. Consider, again, the low-noise amplifier. In general an RF environment will be replete with carriers of other frequencies, all of which will be incident on the receiving antenna. In the case of a perfectly linear LNA, the selectivity of the receiver would be determined solely by the quality of the bandpass filters. If, however, the transfer characteristic of the LNA contains nonlinear terms that are not negligible, intermodulation products could result that fall squarely into the passband of the receiver. Linearity is an issue that cannot be ignored anywhere in the design of an RF front end.

Sometimes a variable attenuator is inserted between the antenna and the LNA. This component warrants a special discussion, which is included at a later point in this chapter. For now, it is sufficient to understand that the input power from the antenna routinely varies over four orders of magnitude [4]. This far surpasses the dynamic range of a typical LNA, and the variable attenuator helps to mitigate the effects of these fluctuations.

There is one more component in figure 1-1 that has not yet been discussed: the automatic gain control (AGC). The AGC usually the last analog component before the analog-to-digital converter of the baseband processing section. Many of the modern wireless protocols carry digital information in the phase of the carrier. Accordingly, the amplitude of the carrier provides no information, and the designer is as liberty to set it to whatever is convenient. "Whatever is convenient" is highly dependent on the particular A/D; if the entire input range of the A/D is utilized, the quantization noise inherent in the digitization process has minimum effect. Thus, an AGC is a very necessary part of a modern microwave receiver.

The RF Environment

The mobile and indoor RF environment has been referred to in this report as "extremely dynamic." As background for understanding this project, it is a characterization that bears considerable clarification. This clarification is the object of the rest of this section. Briefly,
"extremely dynamic" may be understood to be the following: a single antenna will experience "fades" in received power that can be as great as 40dB below the mean, with a typical duration on the order of milliseconds [5].

In a broad sense, the fact that we have to deal with the phenomenon of fading can be attributed to the FCC's allocation of the radio frequency spectrum. Indoor and mobile wireless communications have been assigned portions of the spectrum that begin at frequencies on the order of a gigahertz. The wavelength in free space of a 1 GHz carrier is \( \nu = c/\lambda \) 30 cm. This wavelength is sufficiently short that everyday objects, such as a person walking in front of the antenna, have a significant impact on the propagation of the wave. Consider a single antenna element radiating isotropically in a cluttered room. Walls, books, people, furniture and any number of other objects will produce innumerable reflections of the high frequency wave. The result is a three dimensional standing wave pattern in the room [5]. If we now imagine walking around the room equipped with some sort of electric field strength meter, we will notice substantial variations from point to point. This phenomenon is referred to as fading, and it is the natural way for a standing wave pattern to manifest itself. If we vary our walking speed, we see that the average duration of the fades, and the rate at which they occur, varies significantly. This serves to further bolster our intuition, and provides a good mental picture to keep in mind when thinking about a fading environment.

Within the context of wireless communications, this concept of multiple arrivals at a single point in space is given a special name: multipath propagation. If reliable wireless communications are to be established, it is clear that this type of propagation, and its consequent fading, warrant a thorough characterization effort. When one considers, however, the infinitude of possible configurations of a room or outdoor environment, one is hard-pressed to come up with a general, analytical way to model fading. Indeed, even given a relatively simple room with everything fixed in place, the true standing wave pattern could only be approximated using a computation-intensive simulation algorithm. The widely accepted solution is to employ probabilistic models, which are described briefly here.

Consider a single transmitted signal that is vertically polarized (the following development closely follows that given in [6]). Examining the electric field at any given point in space, we can write it as
\[ E_z = E_0 \sum_{n=1}^{N} C_n \cos(\omega_c t + \theta_n) \]

where
\[ \theta_n = \omega_n t + \phi_n. \]

In these equations, \( \omega_c \) is the carrier frequency, and \( E_0 C_n \) is the amplitude of the \( n \)th reflected wave. We normalize \( \left( \sum_{n=1}^{N} C_n^2 \right) = 1 \) as a matter of book keeping: to do otherwise would be to implicitly assume that fading was at least partially attributable to energy loss, not destructive interference. While that may be true in the real world, the reader should note that we are not considering it here. The \( \phi_n \) are random, uniformly distributed phase angles. In a mobile environment, there is also a Doppler shift to consider; this is the purpose of the \( \omega_n \) terms. Notice that these Doppler terms are necessarily probabilistic. The reason for this is that the magnitude of the Doppler shift is related to the angle between the path of propagation of the \( n \)th reflected wave and the velocity vector of the receiving antenna. We don't know this angle beforehand; it is therefore fitting that \( \omega_n \) be a random variable.

Using a simple trigonometric identity, it is useful to rewrite (1-7) as
\[ E_z = T_c(t) \cos \omega_c t - T_s(t) \sin \omega_c t, \]

where
\[ T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(\omega_n t + \phi_n), \]
and
\[ T_s(t) = E_0 \sum_{n=1}^{N} C_n \sin(\omega_n t + \phi_n). \]

At first it would appear that this does not help us: we agreed that \( \omega_n \) and \( \phi_n \) were random variables, but have so far failed provide any insight on the behavior of the \( T_c(t) \) and \( T_s(t) \). The prospect of a rigorous analysis of (1-10) and (1-11) looks pretty fearsome, until we realize that we are saved by the central limit theorem. \( T_c \) and \( T_s \) are, for fixed \( t \), random variables that are themselves the sum of a large number of random variables. If \( N \) is large enough, which we
conveniently assume, $T_c$ and $T_s$ become Gaussian. They are zero mean\(^1\), and their variance is straightforward to calculate:

$$\langle T_c^2 \rangle = \langle T_s^2 \rangle = \frac{E^2}{2}. \quad (1-12)$$

$T_c$ and $T_s$ are uncorrelated (and, because they are zero mean, independent):

$$\langle T_c T_s \rangle = 0. \quad (1-13)$$

Armed with the knowledge that $T_c$ and $T_s$ are independent, Gaussian random variables, we can go one step further and determine the probability density function (pdf) of the envelope of the carrier. The pdf for a Gaussian variable is

$$P(x) = \frac{1}{\sqrt{2\pi b}} e^{-x^2/2b}, \quad (1-14)$$

where $b$ is the variance calculated in (1-12). The envelope of $E_c$ is expressed as

$$r = (T_c^2 + T_s^2)^{1/2}. \quad (1-15)$$

It turns out that the pdf for the envelope of the carrier is

$$P(r) = \frac{r}{b} e^{-r^2/2b}, \quad r \geq 0 \quad (1-16)$$

$$= 0 \quad \text{O.W.}$$

Far from obvious, perhaps, but proven in [16]. There remains only the phase of the carrier, which is a random variable uniformly distributed between 0 and $2\pi$ [7]. Taking this into account, the envelope is what we call a Rayleigh faded parameter; the corresponding RF environment is called a Rayleigh fading environment. Rayleigh fading accounts for rapid fluctuations in the signal level, and is responsible for the most dramatic fades.

\(^{\text{1}}\) This may not be immediately clear, given that we’ve said nothing about the form of the pdf for $\omega$. But if we accept that $\phi_a$ and $\omega$ are independent, then their joint pdf is $P_a(\phi)P_\omega(\omega)$. The expected value of each cosine term can be expressed as $\int \cos(\omega \tau + \phi_a) P_a(\phi) P_\omega(\omega) d\phi d\omega$. If we integrate over $\phi$ first,
Looking back over the preceding derivation, one might notice what appears to be a flaw in the reasoning behind the model. This flaw would be based on the following argument: *The envelope should be probabilistic, but not to this extent. This model ignores the dominant contribution to the signal at the receiving antenna, which is the direct path. It must dominate, because it loses no energy when interacting with imperfectly reflecting obstacles, and it is necessarily deterministic. This model therefore represents the absolute worst case.* This argument is not completely valid, but it does bring up an important point. In a Rayleigh fading environment, it is assumed that there is no direct path of propagation between the transmitting and receiving antennas. But in a mobile or even indoor area, a direct path of propagation is by no means guaranteed. In fact, upon deeper reflection it becomes conceivable that the existence of a direct path might be difficult to ensure. The model is, then, a good one, and widely applied in the engineering of wireless systems.

Of course, the case of an existing direct path cannot be completely ignored. Necessity has spawned another widely used model, which incorporates the line-of-sight component that was missed in the above paragraph. In short, the received signal is the sum of a deterministic component and a Rayleigh faded component. This is called a Rician fading channel, and the pdf for the envelope of the carrier is [7]

\[
P(r) = \frac{r}{\sigma_r^2} e^{-\left(r^2 + a^2\right)/2\sigma_r^2} I_0\left(\frac{ar}{\sigma_r^2}\right),
\]

(1-17)

where \(a\) is the amplitude of the line-of-sight signal and \(I_0\) is a zero order modified Bessel function.

There is one more type of fading that is observed in mobile radio channels, and it is referred to as log-normal fading [8]. It gets its name from the fact that the pdf of the received power, in dBm, fits a normal distribution. It is primarily discussed within the context of a mobile environment, and is attributed to changes in the surrounding terrain. It tends to cause slower

which we are at liberty to do, we see that the expected value of the cosine, and therefore the expected value of the aggregate, is zero.
fluctuations than Rayleigh fading; so much slower, in fact, that on a plot of signal strength versus distance the two types of fading can be separated visually. Log-normal fading looks like a kind of slowly varying DC offset on such a plot.

We are still missing important information concerning the dynamic nature of the RF environment. If a microwave engineer faces the task of designing receivers that can automatically respond to sudden changes in signal level, his/her primary concern is knowing the time scales over which these fades occur. Should the system be able to respond within a millisecond, or are sub-microsecond response times required? Two widely recognized statistics are used in this connection: the level-crossing rate and the average duration of a fade [5, 6, 8, 9]. The former is the expected rate at which the carrier envelope crosses a given signal level in the positive direction. The latter is the expected time the signal is expected to stay below a certain level. Derivations for both can be found in [6]; in practice, though, numbers are calculated with the aid of such charts as can be found in [9]. Shown below is a table with enough typical numbers to give the reader a feel for what has to be dealt with. These numbers are for a car traveling at 15 mph, receiving a 1 GHz carrier.

<table>
<thead>
<tr>
<th>Fading Level (Electric Field)</th>
<th>Level Crossing Rate (crossings/sec)</th>
<th>Average Duration of Fade (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3 dB</td>
<td>27.8</td>
<td>18.0</td>
</tr>
<tr>
<td>-6 dB</td>
<td>22.3</td>
<td>12.6</td>
</tr>
<tr>
<td>-10 dB</td>
<td>16.7</td>
<td>5.4</td>
</tr>
<tr>
<td>-20 dB</td>
<td>5.0</td>
<td>1.8</td>
</tr>
<tr>
<td>-25 dB</td>
<td>3.1</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 1-1: Time scales for fading events

It is useful to note that the level crossing rates scale linearly with \(V/\lambda\), the ratio of vehicle speed to the free space wavelength; the average duration of a fade scales linearly with the inverse of this ratio. If we keep in mind the picture of a large, complicated standing wave pattern, this result is intuitively pleasing.

Informative though these figures are, what really concerns us is the spectral content of the envelope (e.g., "Is it bandlimited?"). Jakes [6] performs a derivation of the envelope spectrum that far exceeds the rigor required here; the important thing is to know that the fading envelope is bandlimited to twice the maximum Doppler shift, or \(2V/\lambda\). For the aforementioned
car the highest frequency component in the envelope would be 44.7 Hz. Looking at the table, this is close to what you might guess.

The Impact of the RF Environment on Receiver Design

Having thus undertaken to gain a better understanding of microwave propagation, we can now fully appreciate some of the special demands that are made of a modern receiver. The difficulties, while numerous, can be broadly classified into two categories: fundamental information recovery problems, and hardware system issues. The purpose of this section is to provide a brief overview of each type.

The most basic information recovery problem that can occur in a fading environment is a wildly fluctuating signal-to-noise ratio, or SNR. Noise in the atmosphere is extremely broadband; the picture of a standing wave for the noise is thus rendered useless. The noise generated by the receiver circuitry is also broadband (which is to say, "white"). The consequence of all this is that only the signal "fades" in a fading environment; the noise floor does not move. A deep fade can therefore bury the signal under the noise floor, and when this happens the information it carried is fundamentally lost.

Even assuming that the signal stays above the noise floor, we find that there are other difficulties. We have already considered some of the effects of multipath propagation, but we have not considered the case wherein one path is a several wavelengths longer than the other. The signal distortion that results is referred to as delay spread [10]. One can easily imagine the impulse response of such a channel; it could be similar to that shown in figure 1-3, for example. This type of distortion can cause considerable problems in digital communications, where a bit may leave the transmitter as a sharply defined pulse. If the bits are too close together (in time) on a channel with severe delay spread, successive bits could begin interfering with each other. This effect is referred to as intersymbol interference (ISI). It is seen that this characteristic of a transmission channel could place a hard upper bound on the possible bit rate.
Even if the bit rate is below the theoretical upper bound set by the presence of delay spread, ISI can still cause problems in that the bit-error rate (BER) may be higher than it has to be. One way to reduce the effects of the channel is to introduce an equalizer in the baseband digital signal processing. A tapped delay line, such as that shown in figure 1-4a, can be designed to emulate as closely as possible the inverse of the channel transfer function. The improved impulse response is shown in figure 1-4b.

![Equalizer diagram](image1)

**Figure 1-4a:** An equalizer

![Effect of equalizer diagram](image2)

**Figure 1-4b:** The effect of well-designed equalizer on the received signal
The phenomenon of delay spread is an example of temporal diversity: there exist in the RF environment multiple replicas of the original signal, spread out in time. If one were to set up two antennas spaced reasonably far apart (at least a half wavelength [11]), one would observe a different kind of diversity. The antennas would receive copies of the same signal, but the fading observed at one location would be uncorrelated with the fading at the other. In other words, there exist multiple replicas of the original signal that are separated from each other spatially; this is referred to as spatial diversity. If we set up a number of antennas to collect these replicas and then add them together, we can expect that sometimes they will add completely in phase, and at other times they will cancel each other out; most of the time, we will see a medium between these two extremes. If we do the combining more purposefully, however, modifying the gain and phase relationships between the antennas, it is clear from the literature that a great deal can be done to diminish the fading in an RF channel [4, 11, 12, 13]. This will be more fully discussed in Chapter 2. For now it is sufficient to understand that a multiple antenna system, properly designed, has an inherent advantage over a single antenna system.

One of the problems already discussed, the noise floor, also falls into the category of a hardware issue. An imperfect low-noise amplifier actually raises the noise floor of the receiver, whereas a perfect one preserves the input SNR, such as it is. A more serious hardware issue is often that of dynamic range. It has been mentioned that in a fading environment the received power can vary by as much as four orders of magnitude. In the face of this difficulty, one possible approach is to choose a gain for the front end that is a compromise: the extremely weak signals will be lost, and the strong ones will cause the front end amplifiers to saturate. The hope is that most of the time the received power will fall somewhere in the middle.

There is another problem here, even if we are fortunate enough to have the signal just barely fit between the extremes of the front end's dynamic range. The large signal characteristics of active devices are nonlinear; the consequence of our design decision then becomes the increased likelihood that unwanted intermodulation products will be present. The classic low-frequency solution to this would be to close a negative feedback loop around the unruly gain element. Feedback loops at such high frequencies are plagued by phase shifts due to delays in the loop transmission, however. This is in addition to the usual parasitics, which are present in force in a multi-gigahertz system. Sometimes local feedback is employed, such as a resistor between the gate and the drain of a FET (or base and collector of a BJT). In the case of FETs,
however, which tend to be favored because of their inherently superior noise characteristics, the low value of the transconductance (relative to that of a BJT) limits the performance of the amplifier [14]. Moreover, source or emitter degeneration degrades the noise performance of the amplifier, discouraging the use of this technique in the first stage of an RF front end.

At present, the typical solution to the dynamic range problem is to place a variable attenuator in the signal path, as shown in figure 1-1. Somewhere in the system the received power is detected, and the attenuation is varied accordingly. One possible implementation using PIN diodes is shown in figure 1-5 [15].

![An electronically variable attenuator](image)

**Figure 1-5:** An electronically variable attenuator

The incremental resistance of the diodes is varied by controlling the DC bias current through them. In so doing, the amount of attenuation is continuously varied. It can be seen that when $D_1$ has a low resistance and $D_2$ is turned off, the device will pass the signal from input to output essentially unmolested. In the other extreme case, with $D_2$ conducting and $D_1$ turned off, the device completely isolates the input from the output. In addition, it provides well-behaved, resistive terminations at its input and output ports.

The circuit shown in figure 1-5 is an example of the "Bridged-T" attenuator topology. Other topologies exist, most notably the "Pi-circuit" and the "T-circuit" topologies that involve only three resistive elements [15]. An electronically variable resistor can also be implemented.
using the conducting channel of a field-effect transistor; this, as well, has been used to implement variable attenuators.

The Proposed System

So far in this report, we have seen a few of the problems faced by the designer of a modern microwave receiver. Chief among those problems is the phenomenon of fading; it has been stated that spatial diversity can be exploited to counteract this. The need for a variable attenuator has been firmly established. The proposed system is an integration of these two concepts. A block diagram is shown in figure 1-6. The idea is to build an automatic gain control in the RF front end of the receiver. A desired power input is given to the system in the form of a DC voltage, and the controller adjusts the phase relationship between the two antennas until that power level is achieved. If the desired power is greater than the total power available, the controller should adjust the phase until the received power is maximized.

![Figure 1-6: The proposed system](image_url)

In addition to providing the benefits of a continuously variable attenuator and of antenna diversity, this system also has a few other features. Among them:
• The antenna diversity does not come at the price of additional RF parts, such as LNA's. Most multiple-antenna systems do the phase manipulation in baseband processing. This approach requires that each antenna have its own LNA, bandpass filter, and mixer. LNA's and mixers in particular can be expensive RF components.

• It is completely modular, in that it could be used as the front end of any receiver. This would not be the case, say, if the controller itself was one of many processes running on the baseband DSP chip. The system would then be inextricably linked to the receiver it was designed for. In the current form, however, it would be marketable as a stand-alone “smart” antenna.

This is the system that is the subject of this thesis.
Chapter 1 References


Chapter 2: The Theory Behind the Method

The method that the proposed system employs is commonly referred to as “equal gain combining.” As the name implies, the signals received from the two antenna elements are equally weighted, with only their phase relationship subject to the demands of the controller. It will be shown in this chapter that this represents a tradeoff between two other combining methods, achieving an effective compromise between the simplicity of one and the superior performance of the other. In addition, the algorithm implemented by the controller is introduced. The details of the controller itself, together with an analysis of its behavior, are the substance of Chapter 3.

Methods of Spatial Combining

The basic problem that deep fades introduce is one of information loss: the signal vanishes into the noise floor, while the bit-error rate, to extend the analogy, goes through the roof. The ideal would be to transmit and receive a signal that does not fade; this is an impossibility in most modern systems. We can improve things, however, by making several copies of the signal available to the receiver, where each copy fades independently of the rest. A
complete information loss then requires a deep, simultaneous fade on all of the received copies, and we find that the laws of probability work on our behalf. This illustrates the simple concept behind diversity techniques, as they are called: the likelihood of such an event clearly decreases with increasing degree of diversity. We will show that significant gains can be made in going from a single copy to even two-fold signal diversity.

Convinced that diversity is a good idea, we still find ourselves far removed from an actual solution. How, for instance, does one intelligently combine the multiple copies to yield the greatest benefit? Or, more fundamentally, how do you get multiple copies in the first place? The most obvious answer to the second question has already been introduced, which is to employ multiple antennas at the receiver site. By spacing them at least a half wavelength apart, one can ensure that the fading behavior of each signal will exhibit small correlation with that of the others. Other existing methods include: (1) transmitting signals sufficiently separated in time; (2) transmitting copies on orthogonal polarizations; (3) the use of directive antennas (transmitting or receiving) that point in widely different directions; (4) transmitting the same signals on sufficiently different carrier frequencies [1]. The use of multiple receiving antennas will be focused on here, as it is the most common method and the method chosen for this thesis project. Within this context we will examine the more pressing question of how to combine the multiple copies in a meaningful way.

**Selection Diversity**

One rather intuitive approach to the combining problem is, strictly speaking, not to combine. In a selection diversity scheme, a continuous comparison (on the basis of signal-to-noise ratio (SNR)) is made between the diversity branches. The SNR is sometimes difficult to measure in practice; often the branch with the largest signal plus noise ($S + N$) is assumed to have the greatest signal content. The branch with the highest SNR is then connected to the rest of the receiver, and this remains the state of affairs until a different branch exhibits the highest SNR. This approach offers the clear advantage of simplicity, while yielding a substantial benefit in terms of reducing the bit-error rate. It is instructive to examine the nature of this benefit mathematically.

Recall the form of the pdf for the magnitude of a Rayleigh faded envelope:
\[ P(r) = \frac{r}{b} e^{-r^2/2b}, \quad r \geq 0 \]  
\[ = 0 \quad \text{O.W.} \] 

For the purposes of a fading discussion, it is useful to consider the probability that the signal strength will fall below a certain value. We will denote the cumulative distribution function as \( P\{r \leq R\} \), which is the probability that the amplitude, \( r \), is less than or equal to \( R \).

\[ P\{r_i \leq R\} = 0 \quad \text{if } R < 0 \]
\[ = \int_0^R \frac{r_i}{b} e^{-r_i^2/2b} \, dr_i \quad \text{O.W.} \]

We are fortunate in that the antiderivative (2-1) is easily done by inspection:

\[ \int_0^R \frac{r_i}{b} e^{-r_i^2/2b} \, dr_i = \left[-e^{-r_i^2/2b}\right]_0^R \]
\[ = 1 - e^{-R^2/2b} \]  

(2-2)

Our real interest, of course, is in the signal-to-noise ratio. If we assume that the noise power is constant, and that that power is independent of the branch, a recasting of the equation into the appropriate form is not difficult. The mean signal power in a branch, averaged over one cycle, is \( r_i^2 / 2 \), and let the mean noise power in a single branch be \( N \). The probability that the local mean SNR is less than some value \( \gamma_s \) is equal to the probability that

\[ \frac{r_i^2}{2N} < \gamma_s \]  
\[ r_i^2 < 2N\gamma_s \]  
\[ r_i < \sqrt{2N\gamma_s} \]  

(2-3)  
(2-4)  
(2-5)

where that last step is possible because all quantities concerned are positive. The probability that the SNR will be less than some value \( \gamma_s \) can now be expressed as
Finally, we recognize b/N as the global mean signal power to mean noise ratio, and rewrite (2-6) as

\[ P\{\gamma_i \leq \gamma_N\} = 1 - e^{-\gamma_i b} \]  \hspace{1cm} (2-6)

where \( \Gamma = \frac{b}{N} \). Now it can be seen why this method of selection diversity is successful. If we assume that the switching mechanism is instantaneous and free of noise, the probability that the receiver will be fed a signal with SNR less than \( \gamma_i \) becomes

\[ P\{\gamma_1, \ldots, \gamma_M \leq \gamma_i\} = \left(1 - e^{-\gamma_i/\Gamma}\right)^M, \]  \hspace{1cm} (2-7)

where \( M \) is the number of receiving elements (or, more generally, the degree of diversity). Since the quantity being raised to the power \( M \) is always less than one, it is clear that we do better as the number of antennas increases. As to seeing how much better, it is instructive to look at a graph.

\[ f(x) \quad g(x) \quad h(x) \quad k(x) \quad l(x) \]

**Figure 2-1:** Probability that a fade of depth \( x \) will occur for various values of \( M \)

In this graph, \( x \) is \( \gamma_i/\Gamma \), in decibels, and \( f, g, h, k, \) and \( l \) are the cumulative distribution functions for various values of \( M \). If we compare probability statistics at the “likelihood of a 20dB fade”
mark, we see that each new antenna element gets us about two orders of magnitude in improved performance.

It is also, for purposes of a later comparison, useful to look at the mean SNR. The end result of this calculation is [1]:

\[ \langle \gamma \rangle = \Gamma \sum_{k=1}^{M} \frac{1}{k} \]  

(2-9)

There are a couple of important things to take away from the preceding exercise. First and foremost, it confirms our intuition that, from an information recovery standpoint, diversity is a good thing. We took for an example the most naive form of combining, and saw dramatic increases in deep-fade performance. Second, it should be noted that it is probably not in the engineer's best interest to push antenna diversity to the extreme, particularly if the design is for a hand-held unit. You can see from the graph that even the addition of a single element results in a significant improvement, with the extra advantage of allowing a relatively simple switching scheme.

**Maximal Ratio Combining [1]**

Selection diversity has thus far been discussed as an intuitive, first iteration, non-optimized way of beating the fading problem. This characterization, while accurate, should not be taken to mean that it is never used: in many mobile units, where cost is of extreme importance, this method finds ready applicability. Still, we have to believe that, cost aside, there is a way to achieve even higher performance. One way, referred to as maximal ratio combining, is presented here.

In a maximal ratio combining system, the various receiving branches are cophased, weighted in some optimal manner, and summed. A general diagram for such a scheme is shown in figure 2-2.
Figure 2-2: A maximal ratio combining system

It turns out that the optimal weighting (for any linear combiner) is to choose each $a_i$ to be proportional to the ratio of signal amplitude to noise power in that branch. That is,

$$a_i \propto \frac{r_i}{N}.$$  \hspace{1cm} (2-10)

The SNR out of the combiner winds up being the sum of the branch SNR’s. As long as none of the branches has an SNR of less than one, we are guaranteed an improvement with each additional branch. The mean SNR of the combined signal is simply expressed:

$$\langle \gamma \rangle = \sum_{i=1}^{M} \langle \gamma_i \rangle = \sum_{i=1}^{M} \Gamma = M\Gamma.$$  \hspace{1cm} (2-11)

Equal Gain Combining [1]

Equal gain combining is something of a compromise between the two methods previously shown. In this method of fading reduction, signals from all of the antenna elements receive the same weight, and are cophased. The mean SNR for this type of combining is

$$\langle \gamma \rangle = \Gamma \left[ 1 + (M - 1) \frac{\pi}{4} \right].$$  \hspace{1cm} (2-12)
It has the benefit of being simpler to implement than maximal ratio combining, while offering performance that is superior to that of a selection diversity scheme.

We are now in a position to compare the three approaches. This is best shown graphically, although it is clear from the equations which methods offer the best performance per added branch of diversity. With maximal ratio combining, the mean SNR increases linearly with the number of receiving elements. This is also true in the case of equal gain combining, although the slope of that line is smaller. Selection diversity is a clear loser in this contest, particularly in systems which employ a high degree of diversity.

Figure 2-3: A comparison of three combining methods

In Figure 2-3, s corresponds to selection diversity, t to maximal ratio combining, and u to equal gain combining. We see that a switch from selection diversity to equal gain combining yields an impressive boost in performance. The same cannot be said for a similar switch from equal gain to maximal ratio combining.

We have now established a context within which to view the proposed system. A reexamination of figure 1-6 reveals that we are implementing a type of equal gain combining. But there is an important difference: in implementing an automatic gain control, we have departed from the typical scheme in that the two signals will not always be cophased. For purposes of a fading discussion we must consider two cases: (1) the desired power is
unattainable in the given environment; (2) a phase relationship between the two signals exists such that, if obtained, the desired power can be met exactly.

The first case establishes an important specification on the design of the controller. If the situation arises in which the desired power is unattainable, the controller must do its best. That is to say, it must maximize the received power if the command level is too high, and it must minimize the received power if the command level is too low. If it is desirable to mimic the behavior of typical equal gain combining schemes, one need only to set the command level to an unrealistically high level. The reader may find the idea of minimizing the received power curious; in theory, complete cancellation is possible if the gains are equal. Two things, however, render this idea useless in practice. First, based on what we know about fading, we can be almost assured that the signals at their respective antenna elements are not of exactly the same amplitude. Second, even if they were of the same amplitude, complete cancellation would necessitate the use of a perfectly lossless phase shifter. Once again, we run into an idea that is useful only as a theoretical abstraction.

The second case is the best possible if our sole objective is to have a combined received signal that does not fade. However, it is important to note that the question of which case is pertinent has, in general, a very time dependent answer. This is particularly true if the fading envelope of the two received signals exhibits unexpectedly correlated behavior: a deep, simultaneous plunge in received power at both elements may cause an ordinarily reasonable power command to be unattainable.

At this point, the reader should recognize the RF AGC as the system-level tradeoff that it is. We are not maximizing the received SNR, which is what we should do from an information recovery standpoint. Rather, we are deciding on an SNR that is "good enough," and making more effective use of our LNA. In so doing, we are introducing a kind of intelligent behavior that is new to the field of "smart" antennas.
The Problem of Designing a Controller

The focus of this section will be on some of the difficulties that the controller in figure 1-6 must overcome. Also, the basic algorithm is introduced.

Specifications for the Controller

The design of the feedback controller is complicated by the fact that the sign of the loop transmission is uncertain. This problem owes its existence to the nonlinear, time-varying transfer function between a phase change and the resultant change in detected power. This is best shown mathematically; consider the sum of two equal-frequency sinusoids that differ only in phase. If we examine the magnitude squared,

\[ |A e^{i\alpha} e^{i\theta} + A e^{i\alpha} e^{i\phi}|^2 = 2A^2 (1 + \cos(\phi - \theta)) \]  \hspace{1cm} (2-13)

we can see the exact nature of the non-linearity. Denoting the phase difference as \( \alpha \) and differentiating, we can linearize this function and look at its behavior for small fluctuations in phase about a set operating point:

\[ \frac{d}{d\alpha} [2A^2 (1 + \cos\alpha)] = -2A^2 \sin\alpha. \]  \hspace{1cm} (2-14)

If the operating point is beyond our control, as it is in this case, the incremental relationship between a phase change and a change in received power cannot even be ascertained to within a minus sign. In a mobile environment, the problem is exacerbated by the random time-varying nature of this relationship: the phase difference between the two received signals (before they are processed) is what determines the operating point, and this changes as the receiver moves.

All that is clear at this point is that a classic, linear feedback loop is not an option. However, we have something to learn from their basic operation. The key to feedback controllers is that they seek to minimize the absolute value of the error. The direction that the input of the plant needs to be “pulled” to correct for, say, an output that is too small, is known a
priori by the system designer; s/he is thus able to ensure that the sign of the loop transmission is correct.

As it happens, our problem has an additional difficulty: if by guessing we happen to get the sign of the loop transmission right initially, it is probable that this would be a temporary state of affairs. The controller must be able to rapidly adapt to this changing sign.

The fading behavior also gives us constraints (or, at least, goals) to shoot for as far as loop speed is concerned. It has been stated that carrier envelope is bandlimited to twice the maximum Doppler shift, which scales linearly with vehicle speed and as the inverse of the carrier wavelength. It is currently common for mobile systems to use carriers that are close to 1 GHz, and a reasonable worst-case speed would be sixty miles per hour; the spectral content of the fading envelope thus has an upper bound of 180 Hz. Prudence demands that this not be the outer limits of the controller's ability; we chose a safety margin of a factor of ten. Ideally, then, the controller should be able to handle the fading of a 10 GHz carrier when the receiver is moving at 60 mph (an envelope bandlimited to about 2 kHz).

**The Basic Algorithm**

The basic behavior of the controller is based on an algorithm that was introduced and analyzed by Trybus and Hamza [2, 3]. It was intended for use in extremum controllers, whose purpose was to operate an industrial process at an extreme value of the process output; it has since been applied to antenna arrays by Denidni and Delisle [4] to maximize received power. In this implementation the controller moves its output in a given direction for a fixed amount of time, continuously monitoring the error. If the magnitude of the error is increasing, the controller changes direction for the next time step. Otherwise, it keeps moving in the same direction. The system will, in the steady state, fluctuate about the desired value. A simplified diagram depicting this behavior is shown in figure 2-4.
Simple though it is, it can be seen that it satisfies, or has the potential to satisfy, all of the specifications that so concerned us in the last section. We have no reason to suppose that the loop bandwidth has some prohibitive limit. And it is clear that, should the incremental gain of the plant suddenly switch sign, the controller would adapt within one system clock cycle. We have, it would appear, every reason to be optimistic.

**Design Issues**

Of course, implementation rarely fails to bring to light a myriad of "subtle" difficulties. Some of these difficulties will be briefly described in the next few paragraphs.

To begin with, there are performance tradeoffs that simply come with the algorithm. It turns out that the tracking ability of the controller is most closely related to the rate of change of its output (the input to the plant) during a time step. A short thought experiment bears this out: if we imagine a ramp input whose slope exceeds that of the controller output (or, more properly, that of the controller output multiplied by the incremental gain of the plant), the error will increase regardless of the direction chosen. The result will be that it will oscillate, mired in indecision, until the slope becomes trackable. The tradeoff is that for a fixed clock speed, the DC steady-state error increases in proportion to the slope during a step. The best solution is to remove the constraint that the clock rate be fixed. Given the finite bandwidth character of any imaginable implementation, however, we must accept that at some point this option will not be available to us.
There is another option, which is to dynamically vary the slope of the controller’s output as the system deems appropriate. In that way, the aforementioned tradeoff can be beaten to a certain extent. It will be seen in the next chapter how this behavior was added, and the benefit that it produced.

Additional difficulties are introduced when you begin to talk about the real, physical plants that this system is being designed to control. By far the most serious concern is that of the plant bandwidth. We have advanced a 2 KHz bandwidth as a reasonable goal for the loop speed. If we were dealing with a linear feedback loop, this would require the plant to have around 2 KHz of bandwidth (perhaps less, if well compensated). It can be seen in our system, however, that acceptable following behavior would demand a clock rate that is an order of magnitude, if not two, greater than the maximum tracking bandwidth. If the controller is going to know how well it is doing in time to make a good decision on the step direction, the bandwidth of the plant must be correspondingly high. That is, to say the least, of non-trivial consequence. One is forced to accept that this algorithm is best suited to the application its creators had in mind: operating a slow industrial process at an extreme of the process output. Asking it to do rapid tracking (or, equivalently, disturbance rejection) is pushing it in ways that it might not willingly go.

Dynamics of the plant notwithstanding, its linearity is of some concern. A highly nonlinear transfer function between the input and output of the plant will introduce distortion into the otherwise linear step the controller is trying to take. Worse is the prospect of said transfer function having more than one extremum: one can easily imagine a scenario in which the controller gets itself “trapped.”

A full appreciation of these issues was gradually acquired through the process of building and debugging the complete prototype. We now turn our attention to the first major building block of that prototype: the controller itself.
Chapter 2 References


Chapter 3: Design and Analysis of the Controller

In the initial stages of this project, a great deal of time was spent on the details of the basic, non-adaptive behavior of the controller. A subsequent literature search established the age of this "new" idea as greater than twenty years. Still, the design of the prototype proved to be a worthy technical challenge, and the adaptive behavior added as an optimization seems to be a true innovation. The operation of the controller was partially described in Chapter 2. The purpose of this chapter is to examine it in its circuit implementation.

Overview of the Circuit

A simplified diagram of the controller, by way of overview, is provided in figure 3-1.
It can be seen in figure 3-1 how a “step” is mechanized. The voltage input to the op-amp integrator can assume one of two values (one positive, the other of the same magnitude but negative), which is then integrated to produce a ramp for the duration of the clock cycle. The current source, $I_{\text{boost}}$, is an optimization that allows the slope of the ramp to be dynamically varied. The error is computed by a simple difference amplifier, and by examining the sign of the error and of its first derivative it is determined whether the magnitude is growing or shrinking (e.g., if the error and its first derivative have the same sign, the magnitude of the error is growing).

Figure 3-1 also shows that, for the initial characterization effort, the controller output was connected directly to the “power detect” input. In so doing, the circuit is made to mimic a unity gain feedback amplifier. By driving the “CMD” input with a sine wave and observing the output, the bandwidth of the loop could be determined. Of course, this is not quite how the loop will be used in a real system. The “CMD” input will be tied to a DC voltage, and the circuit’s responsibility will be to null out the what would otherwise be disturbances in the received power.

---

2 The word ‘bandwidth’ is used somewhat loosely here. It will turn out that the sinusoidal response and the step response are not mathematically coupled in the same way that they are for a linear system. Additionally, the amplitude of a signal is as important as its spectral content in determining whether the system can track it. One could argue that the term ‘bandwidth’ has lost some of its utility. We will continue to use it here, with the understanding that perhaps ‘speed’ is closer to what we mean.
level. It turns out that a feedback loop’s ability to reject a disturbance is completely equivalent to its ability to follow a command. Figure 3-2 shows why.

\[
\frac{Y(s)}{X(s)} = \frac{L(s)}{1 + L(s)}
\]

\[
\frac{Y(s)}{D(s)} = \frac{1}{1 + L(s)}
\]

**Figure 3-2:** Comparison of command and disturbance transfer functions

If we subtract the disturbance transfer function from unity (which gives us a useful figure of merit for disturbance rejection), we arrive at the same expression relating the command signal to the output. Granted, the diagram depicts a linear feedback loop, which is not what this system is. But the principle still holds, as one can readily convince oneself. For the skeptic, an actual experiment was done that bore this out.

Referring again to figure 3-1, it can be seen that the differentiation was carried out by an actual analog differentiator. This in contrast to, say, using a sample-and-hold circuit with a difference amplifier. Initially it was thought that the latter approach would be safer, given the analog differentiator’s fearsome reputation for noise susceptibility. It turned out, however, that the cost of the op-amp was smaller than that of the S/H by about two orders of magnitude; we were thereby persuaded that the noise problems could be overcome.

**Design of the Non-Adaptive Circuit**

Figure 3-3 depicts an early prototype of the controller, before its performance had been optimized by the addition of the \( I_{\text{boost}} \) current source.
The most fundamental decision concerning the design of this prototype was made right at the beginning of the project: was it to be implemented in analog or digital circuitry? The algorithm itself, simple as it is, could withstand the quirks of either implementation. A digital option had been explored by Jen Wei Liang of the Information Systems Laboratory at Stanford University, during his own sojourn with the department from June until September of 1996. Digital schemes automatically receive credit for being more flexible and, in some vague sense, "cleaner," it is therefore necessary to justify the pursuit of an analog solution to this problem. This is an issue that will be returned to at a later point in this chapter. For now, it is sufficient to list the initial motivation, which was threefold:

- The controller has an analog input, an analog output, and between the two ends must make a very simple decision that is easily mechanized using a few inexpensive logic gates. It was thought that the addition of an analog-to-digital converter, a digital-to-analog converter, and a microprocessor would vastly overcomplicate the system as well as be a waste of

---

3 One could observe, of course, that figure 3-3 displays a solution incorporating both. For purposes of this discussion, 'digital' means 'involves a microprocessor or another such chip with a large amount of processing power, relative to that of a logic gate.'
computational power. One should bear in mind that, at the time, the addition of adaptive behavior had not been considered.

- Prototyping and experimenting were made very easy. Simply switching a capacitor or resistor, as opposed to reprogramming an EPROM, was all that was required to observe the effects of changing various controller parameters. This argument loses some of its weight if the tools for such reprogramming are readily available; for this experiment at Bell Laboratories, they were not. They would have had to be ordered, learned about, etc..

- The author just likes analog circuits better.

A somewhat modified position, acquired after actually completing the project, will be related at the end of this chapter.

Returning to the circuit at hand, it was anticipated that the final clock rate would fall somewhere between 0.5 and 1 MHz. The implications of this decision are easily seen by referring to the data sheets for these common National op-amps: 1 MHz is pushing the limits of their gain-bandwidth product. While not an overly serious design constraint, it forces an awareness of the flexible compensation scheme that the LM301 offers. The order of the discussion will progress along the signal path, which begins at the upper left corner of the diagram and proceeds clockwise.

**Difference Amplifier**

But for being the first circuit block in the signal path, there was nothing remarkable about this amplifier. It utilized an externally compensated LM301, with a 26 pF compensating capacitor.

**Error Tracking**

This circuit block includes the op-amp differentiator, the two comparators, the XOR gate, and the inverter, and it immediately follows the difference amplifier. The purpose of this block is to decide whether the controller's output is moving in the right direction. If the
magnitude of the error is increasing, the output of the block is a digital '1' that forces the controller to change direction.

The question of how to monitor the magnitude of the error was an interesting one. Given that we've committed ourselves to an analog implementation, the logical thing to do might be to cascade a folding amplifier and a differentiator. However, this approach suffers in that a folding amplifier with fast enough dynamics is difficult to build with off-the-shelf, inexpensive op-amps. It has to be possible if one spends the time to carefully design one out of discrete transistors, but this option was not explored. The present solution combines the virtues of simplicity and speed, and proved to be up to the task.

The design of the differentiator itself was fairly straightforward, once the difficulties inherent to the circuit had been fully appreciated. One serious concern is that of noise susceptibility: given that random noise tends to have strong high-frequency components, the transfer function of an ideal differentiator is problematic. Another is that of stability. In the inverting configuration, the input-output transfer function can be shown to be

\[
H(s) = -\frac{Z_2}{Z_1 + Z_2} \left( \frac{a(s)}{Z_1} \right) \left( 1 + \frac{Z_1}{Z_1 + Z_2} a(s) \right), \tag{3-1}
\]

where \(Z_1\) is the impedance in the input branch, \(Z_2\) is the impedance in the feedback branch, and \(a(s)\) is the open-loop transfer function of the op-amp. A minor manipulation of (3-1) yields a form in which the ideal behavior can be more clearly seen:

\[
H(s) = -\frac{Z_2}{Z_1 + Z_2} \left[ \left( \frac{Z_1 + Z_2}{Z_1} \right) \frac{a(s)}{Z_1} \left( 1 + \frac{Z_1}{Z_1 + Z_2} a(s) \right) \right] \tag{3-2}
\]

\[
= -\frac{Z_2}{Z_1} \left( \frac{Z_1 + Z_2}{Z_1} \right) \frac{a(s)}{1 + \frac{Z_1}{Z_1 + Z_2} a(s)} \tag{3-3}
\]
We recognize the $-Z_2/Z_1$ as the transfer function of an ideal op-amp connected in the inverting configuration. We can imagine the circuit as a cascade of this ideal block and a unity gain feedback follower; the latter of these two system blocks is represented by the second product term in (3-3). If we do this, the loop transmission is clearly identifiable, and we can proceed to do a stability analysis. Substituting $1/Cs$ for $Z_1$, $R$ for $Z_2$, and $K/s$ for $a(s)$ (any op-amp that uses dominant-pole compensation can be modeled this way):

$$L(s) = \frac{K}{s(RCs + 1)}.$$  \hspace{1cm} (3-4)

Given that a large $K$ can be taken for granted in most modern op-amps, it is clear that we are headed for uncomfortably small phase margins. Our predicament is best seen on a root-locus plot, where the combined presence of large $K$ and unmodeled poles would seem to ensure the presence of singularities in the right-half plane.

For the chosen impedances, the ideal part of the transfer function in (3-3) will be the differentiator that we desire. However, we have shown that this connection has stability problems that cannot be ignored. The solution for this system was to insert a resistor in series with the capacitor in the input branch. This well-known remedy introduces a zero in the loop transmission (3-4) which allows the designer to ensure a non-zero phase margin. Moreover, it introduces a pole in the ideal transfer function, causing the circuit to act like a differentiator only for signals well below the pole frequency. In so doing, we are partially relieved of the high-frequency noise problem.

The component values for this differentiator were picked such that the pole in the ideal transfer function occurred at about 1 MHz, where the gain was approximately 10. A feedforward compensated LM301 was used here; a casual look at the op-amp specifications reveals that the limits of its performance are being pushed. This is not without its advantages: the closed-loop transfer function actually rolls off in the vicinity of the pole instead of staying flat, which helps to further limit the effects of high-frequency noise. This circuit proved to be a bottleneck from a speed standpoint, in that it placed a hard limit on the highest useable clock speed.
The comparators were LM311, and the all of the digital logic was from the 74LS (Low-power Schottky) family.

*Step Direction*

The remaining logic in the circuit acts on the decision of the previous block, which was to either change direction or continue in the same direction. This circuit block includes another XOR gate, a NAND gate, an inverter, and a D flip-flop. In essence, it amounts to a glorified T flip-flop; logic is added to allow temporary forcing of the output in one direction or another. The need for this added capability will be further described in a later section.

*Level Shift*

The output of the last block was a digital signal; a “0” indicated that the output should begin increasing, a “1” the exact opposite. Given that an op-amp integrator was used to generate the ramp output from a fixed voltage level, it was necessary to perform a level shift so that a “0” corresponded a negative voltage. This circuit block, consisting of two LM310 buffers and an LM311 comparator, accomplished this.

The two buffers, in concert with a few passive components, establish the voltages that correspond to the two digital levels. The positive voltage was pretty straightforward to implement, given the fact that a LM311 has an open collector output. The negative voltage was something of a trick: once established with the buffer, the diodes, and the resistor, it was connected to the signal ground of the comparator. When the output transistor of the LM311 saturates, it pulls the output down to this level instead of ground.

Of course, this function could have been implemented using a regular op-amp with its output appropriately (diode) clamped. As it happened, the superior switching speed offered by the LM311 contributed to the efficient operation of the circuit.
Ramp Generation

The main part of this circuit block is, again, unremarkable. It is simply an integrator built around a LF347 op-amp.

The surrounding comparators were not there in the very first prototype. It was observed in this early version that sometimes the output of the controller would lock up and at a supply rail and refuse to budge. One can imagine a countless variety of ways in which the integrating capacitor can inadvertently acquire unexpected charge. The problem lies in the fact that when this happens, the output stage of the op-amp is saturated and the incremental gain of this block drops to almost zero. No matter which direction the controller turns, it is unable to improve the error situation. The result is an oscillation near the supply rail, and we observe the output to have locked up.

The comparators in this final block allow the user to preset upper and lower bounds for the controller output. Should it step outside of these limits, comparators force the direction of the controller until the situation is corrected. The shown resistor values allow a swing from -10 to +10 volts, using a +/- 15 volt power supply.

Clock Circuit

In the original prototype, the clock was generated using a 2 MHz crystal oscillator and a counter to divide it down. This proved difficult to control, however: transitions tended to have excessive ringing. Worse, this ringing was present at virtually every node in the signal path. After a couple of days of trying to deal with this, the crystal oscillator was scrapped in favor of a ring oscillator fashioned from TTL inverters. Again, a counter was used to slow it down for an actual clock circuit. Its performance has, so far, been adequate.

Testing and Analysis of Non-Adaptive Circuit

Having built a prototype of the controller, it was found that there was a lot to be learned from experimenting with it. The references listed at the end of chapter 2 make important
comments concerning the behavior such a system. But it is quickly apparent, particularly in the papers by Trybus and Hamza, that the authors had a different problem in mind when formulating their respective treatises. Specifically, they were not too concerned with the issue of tracking a rapidly moving extremum; this caused their papers to be of limited utility when it came to designing the present system. Hamza does offer a rough estimate: "...it may be said that if the position of the extremum can be considered to be constant within a period of the order of four time constants of the system or more, the extremum can be satisfactorily identified." It was desirable to determine this limit a little more precisely for the present system.

The paper by Denidni and Delisle was useful in that it alleviated some doubt that the antenna problem could be tackled this way. Their system was designed to prove the effectiveness of the algorithm in an indoor environment. Accordingly, speed was not an issue for them, and any discussion concerning the actual amount of time it took to find an extremum is entirely absent. Rates of convergence are instead expressed in terms of "iterations," and the improved fading envelope is presented as a function of distance (from the source) rather than time.

A great deal of time was thus spent characterizing all aspects of the controller's behavior. What follows is a discussion of the controller when it was connected as shown in figure 3-1.

**Behavior in the DC Steady State**

For this experiment, the "CMD" input was connected to a variable DC voltage source. In the case of a DC input, we expect to see a steady, triangle wave oscillation about the desired value whose amplitude is determined by the size of the steps being taken. It is straightforward to determine this expected amplitude. The time domain relationship between the input and output of the integrator block is

\[
 y(t) = \frac{1}{RC} \int_{-\infty}^{t} x(\tau) d\tau,
\]  

(3-5)

where R and C are, respectively, the input resistor and the value of the integrating capacitor. In the present case, the input to the integrator takes on one of two DC values, of the same
magnitude but with differing signs. If we denote this magnitude as $K$, we can readily determine the peak-to-peak amplitude of the oscillation:

$$A = \frac{K}{RT}.$$  \hspace{1cm} (3-6)

In equation (3-6), $T$ is the period of the system clock.

One must be careful to note that this is the best case, or the smallest possible oscillation. Which to say, it assumes perfect (hysteresis-free, offset-free, delay-free) comparators operating on noiseless signals. In a noisy circuit, such as this is bound to be, one expects that a small step size may be undone by the inability of the comparators to deal with small errors. The DC behavior would then be characterized by two (or more) steps in one direction before reversing, with a corresponding increase in the size of the oscillation.\(^4\)

Figure 3-4 shows a photograph of the output, when the input is a DC voltage.

![Figure 3-4: Controller output for DC input; 100 mV/div vert.; 2 μs/div horiz.; 179 KHz.](image)

A great deal of insight can be gained by examining figure 3-4. First, by way of confirming our initial attempt an analysis, we do in fact observe a periodic triangle wave. The clock frequency here was 714 KHz, whose 1.4 μs period dictates a step size of about 112 mV for the given component values. The period of the waveform in figure 3-4 is 5.6 μs, exactly four times the period of the clock. We see that in the steady state the circuit takes two steps before reversing, not one as we had hoped.

\(^4\) The skeptic may point out that this analysis smacks of hindsight; in this case, he is correct. The behavior was first observed, and later recognized for what it must be.
A closer look at figure 3-4 reveals some more disturbing behavior: the extreme left edge of the trace is lower than it should be, and gives the impression that maybe this isn't a steady-state oscillation after all. Figure 3-5 provides a better picture what is going on.

In looking at this picture, it is suddenly clear that we have overlooked another non-ideality: asymmetric step sizes. Unless the "down" and "up" voltages have exactly the same magnitude, the overall controller output will exhibit a drift in the dominating direction. It turns out that the diode voltage references in figure 3-3 differ by 5 mV. Using eq. 3-5, we discover that this discrepancy would cause a drift of .3 mV/μs. While not sufficient to completely account for the drift in figure 3-5 (we have neglected the difference in offsets in the LM310, for instance, which could increase the drift rate by as much as a factor of four), it is clear what is causing this behavior.

**Step Response**

It is not difficult to imagine what the circuit will do in response to a step input: it will take many consecutive steps in the appropriate direction until it reaches the new DC level, and then assume the behavior exhibited in the last section. As a figure of merit, we can establish a "step recovery time" which is the time between the beginning of the step at the input and the achieving of the new DC level. Based on the step size, s, and the clock period, T, we can calculate what this should be:

\[
\tau_{\text{recovery}} = \frac{A}{s} T,
\]

(3-7)

where A is the size (in volts) of the step. The actual step response is shown in figure 3-6.
The formula predicts, based on a 112 mV step size, a step recovery time of 62 \mu s; the figure shows about 80 \mu s. Figure 3-4 actually provides a ready explanation for this, as we can see that two steps does not quite cover 200 mV. Moving backwards from eq. 3-7, we can determine the measured step size to be about 87.5 mV.

**Tracking Speed**

The ability of the controller to track a rapidly varying input is, as it happens, the most important characteristic to determine for this particular system. In light of previous observations (see The Problem of Designing a Controller; Design Issues, Chapter 2), it stands to reason that a constant slope input, or ramp, is the natural choice for considering this aspect of the controller’s behavior. Figure 3-7 shows the response to a triangle wave that is easily tracked by the circuit.

Immediately, we can establish a theoretical upper limit for the slope that the controller will be able to track. If the slope of the input exceeds that of the controller’s output during a step, the circuit will be “tricked” into making a wrong decision and find itself in a situation from which it cannot recover. We have measured the step size of the present circuit to be 87.5 mV over the course of a 1.4 \mu s clock period. This works out to a maximum trackable slope of
Looking at figure 3-8 we can already see signs of trouble: a slight bulge on the descending slope. The slope here is $4.5 \times 10^4$ V/s, which is 28% smaller than our theoretical limit; by way of explanation, we retreat again to the shelter of a non-ideality. In the case of figure 3-8, we can see that the derivative of the error at any given time, to within a minus sign, is $(6.25-4.5= 1.75) \times 10^4$ V/s. The corresponding output of the differentiator is $(1.4K\Omega*1nF)\times 1.75\times 10^4$ V/s, or 25 mV. For solderless breadboard projects with average power supplies, one suspects that this is
uncomfortably close to the noise floor. Figure 3-12, which shows the trace when the oscilloscope probe is connected to the ground strip, would seem to bear this out.

Figure 3-12: The signal at system ground; 50 mV/div vert., 0.05 μs/div horiz.

This noise component suggests that as we move just beyond the triangle wave “cutoff” frequency, we can expect the controller’s decisions to take on an increasingly random nature. When we greatly exceed this cutoff, it turns out that these decisions continue to be random even though the derivative of the error at any given time has re-emerged from the noise floor. In this situation, it is equally likely that the controller will perceive an improvement or a degradation as a result of a decision. The end effect is that the output of the controller could probably be modeled as a random variable with a uniform pdf. Figure 3-13, illustrates this case.

Figure 3-13: Greatly exceeding the triangle wave cutoff frequency; 5 V/div vert., 50 μs/div horiz.

**Sinusoidal Bandwidth**

Having learned a great deal from observing its DC steady state behavior, transient response and tracking speed, we are now in a position to better understand what is going on when the device tries to track a sine wave. But first the question must be asked: why would this [further investigation] be valuable? After all, one could argue, we have already discovered the most fundamental aspects of its behavior. The answer is threefold. First, such a test proceeds out of a bias that one cannot avoid inheriting if one experiments with linear systems for too long. Second, a sine wave is novel in that its first derivative is continually varying. While this is no serious argument that a sine wave is a fundamentally new stimulus, it is sufficient to make an
experimenter curious. Third, it will turn out that we still have more to learn about this nonlinear controller.

We begin here the same way we began in the last section, which was to establish some theoretical upper bound on the system bandwidth. One logical course would be to identify the maximum slope that occurs during a sine wave, and based on the system’s demonstrated tracking ability establish a bandwidth limit. From the last section, we know that $4.5 \times 10^4 \text{ V/s}$ is the maximum tracking speed, and for a $5 \text{ V}_{pp}$ sine wave we can determine the corresponding frequency:

$$slop_{\text{max}} = A\omega = A(2\pi f) \quad (3-8)$$

$$f = \frac{slop_{\text{max}}}{2\pi A} = \frac{45000\text{V/s}}{2\pi(2.5\text{V})} = 2.86\text{KHz}. \quad (3-9)$$

The next few figures show the circuit’s attempts to track a sine wave input.

**Figure 3-14**: Sine wave; $5 \text{ V}_{pp}$, 645 Hz.

**Figure 3-15**: Sine wave; $5 \text{ V}_{pp}$, 2.86 KHz.

**Figure 3-16**: Sine wave; $5 \text{ V}_{pp}$, 3.45 KHz.
Figure 3-17: Sine wave; 5 V<sub>pp</sub>, 5 KHz.

Figure 3-18: Sine wave; 5 V<sub>pp</sub>, 11.4 KHz.

Two things stand out in this sequence of photographs. The first is that we were satisfyingly close in our prediction of the sinusoidal bandwidth. The second is that, as we exceed this bandwidth, we don’t observe the near-apocalyptic failure that we saw in the case of a ramp stimulus. Instead, the controller’s output begins to look like a square wave with an amplitude that decreases with increasing frequency (remarkably like a single pole roll-off, as it happens).

This interesting behavior is not difficult to explain. The most fundamental observation we can make is that regardless of frequency, any sine wave will have two regions in its cycle wherein it is momentarily trackable. These regions are centered about the zero-crossings of the first derivative. While we are within the controller’s sinusoidal bandwidth, the width of these regions are such that they encompass the entire waveform; outside of this bandwidth, they shrink as the frequency increases. The result is that the system is periodically given a chance to regain its bearings, so to speak, and thereby avoid a descent into the total chaos of figure 3-13.

Summary

Experiments such as those carried out in the preceding discussion were valuable in that the understanding they imparted greatly aided the optimization effort. It was vital to understand that tracking speed was most closely related to the slope of the controller’s output during a step, as well as the limitations on the system clock speed. It was also important to see that a fixed clock speed forced a tradeoff between tracking speed and DC steady state performance.
These experiments also revealed a comforting feature of the algorithm itself: it is robust to various non-idealities that were not fully considered before building the prototype. Chief among these was the lack of symmetry in the “up” and “down” step sizes; automatically dealing with the noise floor was another. Such behavior was cause for considerable optimism that the whole system could be made to work.

Adaptive Behavior, Part I: Fast Transient Response

In the engineering of linear feedback systems, a designer often faces tradeoffs similar to those that have been discussed here. Low frequency performance must sometimes be sacrificed for high frequency performance and vice versa; all too often it is impossible to optimize both. Here we have an advantage in that we are considering a clocked, nonlinear system: in theory it should be possible to dynamically vary controller parameters. This would allow the system to adapt to different circumstances. That said, the task of deciding how to do this intelligently remains a difficult problem.

What parameters can we vary? We really only have two choices: clock speed, and the rate of change during a step. The former makes almost no sense for two reasons. The first is that, as we saw in the experiments with the basic controller, clock speed has nothing to do with determining the maximum trackable slope of the system. Second, a fast clock keeps the output tightly focused around the desired value regardless of the speed of the input. It stands to reason, then, that the design engineer is running the clock as fast as possible; no motivation can be found for changing that. We are left with varying the rate of change during a step.

To test this idea, it made sense to build into the prototype an ability to “shift gears” in the case of an abrupt transient (such as a step) input. An as yet undesigned transient detect circuit would produce a digital signal indicating that the controller needed a larger step size, and some additional circuitry would make it so. This “additional circuitry” was designed around the realization that the step size could be varied by the addition of the $I_{\text{boost}}$ current source in figure 3-1. The magnitude would be pre-set, and the direction of current flow would be controlled by the same direction signal as in the main circuit. Figure 3-19 shows the circuit to accomplish this.
When the transient detect signal is active (low, in this case), it is seen that the boost current available to the circuit is suddenly made non-zero. The question of how to create a bi-directional current source here proved to be interesting; figure 3-19 betrays the designer’s familiarity with common op-amp design. The core of this “direction changer” is a current source loaded emitter-coupled pair, which alternately allows current to be pushed out of or pulled into the output node depending on the differential voltage between the two bases. An easy mistake to make in implementing this circuit is to leave off the pnp transistors that provide the level shift for this differential pair; the output of this current source is connected to a virtual ground, so TTL voltages applied to the base of the npn’s would cause them to saturate.

Having solved the problem of how to shift gears, the problem of how to tell when to shift remained. While it wasn’t clear at the outset how this should be done, it was clear how it should not be done: the decision should not be based on the nature of the command input. This is a particularly seductive error, given the nature of the experiments that have been performed so far. But in the final system this input is likely to be a fixed voltage reference; it cannot provide any
information about the disturbance (fading) the system is trying to reject. We must come up with another solution.

The solution for this system was based on the following realization: in the case of an abrupt step in the command or disturbance, the controller will respond by taking several consecutive steps in the same direction. A transient detector could be built around a simple counter; the circuit would activate the ‘transient detect’ signal when a predetermined number of steps in the same direction had been taken. Such a circuit is shown in figure 3-20.

![Figure 3-20: Fast transient detect.](image)

It can be seen that the clear input on the counter is asserted whenever the controller changes direction; the “change direction” signal is taken from the error tracking block in the main controller, just after the XOR gate and just before the inverter. Otherwise, the counter increments with each rising edge of the clock. In the DC steady state, the highest number the counter will ever attain is ‘1.’ If the count ever reaches five, the NAND gate disables the counter, holding it at five. It also asserts the transient detect output.

For a square wave at the command input, we can now expect the system to respond by taking five steps in the appropriate direction, changing the step size and racing to the new level, and then shifting back to the original step size. But the circuit in figure 3-20 has an additional behavior: the transistor, 1.25 MΩ resistor, and 681 pF capacitor act as a crude “one-shot,” leaving the transient detect signal asserted for a short time after a direction change takes place. It turned out that a fast sine wave (near the edge of the basic controller’s bandwidth) would trigger
the transient detect. By including this delay, the new step size could be maintained between zero crossings, with the effect that the sinusoidal bandwidth increased as well. Figures 3-21 and 3-22 show the improvement in step recovery time. It is particularly easy to see in figure 3-22 how the controller adjusts its step size back to the DC steady state value.

Figure 3-21: 5V step, 20 µs recovery time.

Figure 3-22: 5V step, magnified to show step size adaptation.

Figure 3-21 shows that we have gained approximately a factor of 4 in step recovery time. It turns out that we gain about a factor of 4 in the sinusoidal bandwidth as well, as figures 3-23, 3-24, and 3-25 show.

Figure 3-23: 1.9 KHz, 5 V_{pp} sine wave.

Figure 3-24: 6.5 KHz, 5 V_{pp} sine wave.
Figure 3-25: 11.1 KHz, 5 V<sub>p-p</sub> sine wave. This is the fastest sinusoid that the controller can now track.

Figure 3-24 is a little disturbing. It looks like the controller is staying on track most of the time, but sometimes it reverts to the square wave behavior that we saw when we exceeded the bandwidth of the basic controller. This turns out to be a timing issue with the one-shot: if the time constant is increased, this problem goes away. It is here that we encounter yet another tradeoff. If the time constant is too long, the settling time for a square wave becomes intolerable.

The shortcoming of this new system, exposed in figure 3-24, is suggestive of more sophisticated adaptive behavior. The fundamental problem is that for the broad range of frequencies from 3 KHz to about 10 KHz, neither of the controller’s step sizes are truly appropriate. Unable to resolve this difficulty, the system seems to try a compromise, wherein perhaps the time-averaged step size works out to be suitable. This behavior prompted a consideration of whether it were possible to equip the system with a continuum of step sizes. Failing that, perhaps a few intermediate step sizes could be added to relieve the awkwardness of the step size hand-offs.

Adaptive Behavior, Part II: Adaptation on a Longer Time Scale

The issue now is slightly different than it was in the last section. Before, we were interested in increasing the system bandwidth without changing the DC behavior. Now the bandwidth that we want is basically there; we just have to overcome some difficulties in the controller’s new mid-range.

The fact that the controller can basically perform well in this frequency range suggests a strategy for overcoming this final hurdle: we can monitor the controller output to find out what kind of signal it is trying to track. This would not be true if, instead of struggling valiantly
through this awkward mid-range, the controller simply did not respond until we got up to around 10 KHz or so.

A workable solution is described as follows. The controller output continues to be connected as before, but it is also connected to the input of a new slow adaptive circuit. The signal is low-pass filtered remove the high-frequency “fuzz” on the wave form (there as a natural consequence of the basic controller algorithm), and then passed through a differentiator. The output of the differentiator goes to a peak detector with an appropriately chosen time constant. In so doing, we have created a DC signal proportional to the highest slope that periodically occurs in either the command or the disturbance wave form. A current proportional to this voltage is generated, and it contributes to the magnitude of $I_{boost}$. The final circuit is shown in figure 3-26.

![Figure 3-26: Slow adaptation circuit.](image)

The circuit shares its output node with the fast transient detect current switch in figure 3-19; notice, however, that the two adaptive mechanisms never contribute current simultaneously. When a fast transient is detected (transient detect out goes low), the current from the slow
adaptation circuit is diverted to ground. It is desirable to decouple the two mechanisms in this manner. Otherwise, inordinately large step sizes result.

After the voltage buffer, the first op-amp in the signal path is the core of a Sallen-Key low-pass filter. The purpose of this filter is to strip off the 179 KHz oscillation that we saw in figure 3-4. This oscillation is clearly of no interest\(^5\): we are trying to look at signals within the sinusoidal bandwidth of the controller. Accordingly, the cut-off frequency of this filter is at roughly 45 KHz.

Next in the signal path is differentiator. In light of the previous discussion of this topology, little need be said. The pole in its ideal closed-loop transfer function was placed at 27 KHz.

The peak detector was built around a “super diode,” a circuit which uses the high gain of the op-amp to nullify the .6 V voltage drop of the 1N914 diode. The time constant of the RC network is about 5 ms, which was experimentally determined to be suitable. This value turns out not to be very critical, as long as it is greater than a millisecond or so.

The final op-amp, together with an npn transistor, functions as a transconductance amplifier. The low pass filtering here provides a final measure of protection against sudden transient spikes that were observed; these were probably due to the occasional activation of the fast transient detect circuit. Figures 3-27, 3-28, 3-29, 3-30, and 3-31 show the improvement in performance that resulted from adding this slow adaptation circuitry.

\(^5\) Indeed, for the sake of stability it is vital that this oscillation be greatly attenuated. In adding this circuit, we have created a second feedback loop in the system: one can imagine a kind of “slew-rate runaway” if the designer is not careful. Here, the loop in question is well beyond crossover at the system clock frequency.
Figure 3-27: 2 KHz, 5 V_{p-p} sine wave.

Figure 3-28: 4 KHz, 5 V_{p-p} sine wave.

Figure 3-29: 6.5 KHz, 5 V_{p-p} sine wave.

Figure 3-30: 8 KHz, 5 V_{p-p} sine wave.

Figure 3-31: 10 KHz, 5 V_{p-p} sine wave.

It is evident that, while we have not completely solved the problem, we have made a great deal of progress. In particular, it is worth comparing figure 3-29 to figure 3-24.
Analog Processing vs. Digital Processing

By way of summary, a block diagram of the system with the improved controller is shown in figure 3-32.

![Block diagram of final controller.](image)

Figure 3-32: Block diagram of final controller.

A decision was made at the outset to design a primarily analog solution; the reasons for this have already been mentioned. But the thought of giving the controller the ability to adjust itself, using only sequential logic and analog circuits, necessarily gives us pause: would this not be a good time to reconsider our exclusion of advanced processing hardware? Perhaps. Regardless, it remained that an exploration of what could be done with simple processing blocks had not been carried out. The knowledge gained from such a study would almost certainly benefit the engineer whose task it was to implement this system with a microprocessor.

For the duration of the project, however, the question of analog vs. digital continued to be raised by many to whom this work was described. This was mostly out of reflex, rather than careful consideration, and in some cases symptomatic of a general distaste for things analog. Be that as it may, there are at least two situations in which the superiority of a microprocessor-based implementation cannot be denied:

[1] A truly general, “plug and play” controller that could be placed in almost any situation would clearly have to be microprocessor driven. One could imagine a kind of autocalibration sequence, wherein the entire transfer function of the plant was mapped out and stored in system memory. Based on this information, an optimum combination of slew rate and clock
speed could be chosen for various operating points along the nonlinear transfer characteristic.

[2] Even if we were to stick with the present controller, it may be that more sophisticated behavior would demand the services of a microprocessor. It is probably true that greater sophistication could yet be achieved using simple decisions and analog measurements; nevertheless, one suspects that after many more design iterations we would be unable to lay claim to the virtue of simplicity.

The importance of the work described in this chapter is that it explores, in a way not seen in the literature, the dynamic behavior of this simple, robust algorithm. Based on this knowledge, two optimizations were developed that dramatically improved performance. If this work is continued by others, it may turn out that the analog implementation is forsaken in favor of a microprocessor. At the very least, then, this circuit stands as a novel solution to an already interesting control problem.
Chapter 4: Completing the RF Prototype

To this point, the controller showed considerable promise. It stood admirably as a proof-of-concept: it was immune to sudden changes in the sign of the loop transmission and, based on the development in Chapter 2, the loop bandwidth seemed sufficient. Still, as a demonstration it was incomplete for two reasons. First, the difficulties associated with implementing a working RF system remained unexplored. Second, the performance of the controller when faced with a truly nonlinear plant had not yet been studied. It was desirable to pursue the completion of a full RF prototype.

This resolution necessitated the design of an analog phase shifter and of a power detector. Chapter 4 covers the design of these two components in less detail than in the case of the controller, as their development was not the focus of this research.

Overview of the RF system; Design of the Phase Shifter

Figure 4-1 depicts the placement of the new components in the overall system (see also figure 1-6).
Design constraints for the phase shifter were chosen with an eye toward speed of development. Simplicity was foremost among these constraints: common, off-the-shelf parts (and not very many, at that) were to be used, and then only if a reasonably priced unit could not be purchased. The phase relationship would be controlled by an analog voltage; the op-amps used constrained this voltage to a range of $\pm 15$ volts. A range of $\pm 5$ volts was more desirable, however, as it would limit the impact of large-signal dynamics on system behavior. In response to a voltage in the established range, the phase shift would vary over a full $360^\circ$ range. Finally, speed of response was critical. The system clock rate is limited by the slowest component in the controller's signal path; we would do well to avoid lowering that limit with the addition of the phase shifter.

A brief survey of the literature yielded a suitable circuit topology [1]. Composed principally of a quadrature hybrid coupler, two varactor diodes, and two inductors, the circuit depicted in figure 4-2 met all of the aforementioned design constraints.
This circuit will introduce a continuously variable phase shift over a 180 degree range; for this project, two were built and cascaded.

Central to an explanation of this circuit is a description of the quadrature hybrid coupler. It is a reciprocal, four-port network. Power incident on port 1 splits evenly between ports 2 and 3, with a 90° phase lag in the former and a 180° phase lag in the latter. If ports 2 and 3 are properly terminated (i.e., fitted with a termination that causes no reflections), no power is detected at port 4. In figure 4-2, indices have been organized such that power incident on any one port is split evenly between the opposite pair, with the diagonal port characterized by the 180° phase shift. Mathematically, the properties of this network can be expressed as the scattering matrix [2]

\[
[S] = -\frac{1}{\sqrt{2}} \begin{bmatrix}
0 & j & 1 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & j \\
0 & 1 & j & 0
\end{bmatrix}
\]

Figure 4-2 shows RF power incident on port 1, but with ports 2 and 3 “improperly” terminated. Indeed, they are terminated in the worst possible way: using only reactive components, no power can be dissipated and thus all must be reflected. The key is that the reflections occur with a phase shift, determined by the value of the terminating impedance. The impedance is easily expressed:

\[
Z = \frac{1 - \omega^2 LC}{j\omega C}.
\]
From here we can determine the reflection coefficient \([3]\), a complex constant whose magnitude is necessarily unity:

\[
\Gamma = \frac{Z/Z_o - 1}{Z/Z_o + 1},
\]

(4-3)

\[
\Gamma = \frac{1 - \omega^2 LC - j\omega CZ_o}{1 - \omega^2 LC + j\omega CZ_o},
\]

(4-4)

where \(Z_o\) is the characteristic impedance of the interconnecting transmission lines. Equation 4-4 confirms our assertion that the magnitude of \(\Gamma\) is unity: it is easily seen that the numerator and denominator are complex conjugates of each other. A look at the limiting behavior \((\omega=0; \omega=\infty)\) reveals the 180° phase range that, at least, makes it seem as though we’re headed in the right direction. All that remains is to show how these terminations and the hybrid coupler combine their functions to yield a complete phase shifter.

Two things should be understood at the outset. First, the varactor capacitances are varied together: their capacitances are always equal to each other. Second, the limiting behavior analysis described in the last paragraph bears some modification: we will be varying the capacitance, not the frequency, and the range of available capacitances is far from infinite. The series combination of the varactor, adjusted to the middle of its capacitance range, and the inductor should therefore resonate at the operating frequency of the incident RF. This will ensure that whatever the range of the varactor, the range of available phase shifts will be maximized.

Now, consider again the circuit shown in figure 4-2. The signal incident on port 1 has been split between ports 2 and 3; port 2 has a -90° phase shift, port 3 has a -180° phase shift. The resultant reflections will combine in some manner at ports 1 and 4, and we can now consider these separately. Port 2’s contribution to the signal at port 4 will have a total phase shift of -270° + \(\phi\): it picks up \(\phi\) from the reflective termination, and another -180° on the second trip through the coupler. Port 3’s contribution will also be -270° + \(\phi\), and we see that we have accomplished our purpose. At port 1, it turns out that the reflections from ports 2 and 3 are 180° out of phase, and thus completely cancel.
Figure 4-3 depicts the phase shifter used in the prototype.

![Diagram of phase shifter](image)

*Varactor diodes: Alpha Industries SMV1200-49 high ratio hyperabrupt junction varactors. Capacitance range: 2.8-28 pF.

**Figure 4-3**: Circuit used to achieve up to 360° phase shifts in prototype system. The RF input was assumed to be a 250 MHz carrier. Note also that the control voltage is from 0 to 5 volts.

The necessity of providing each varactor diode with the same bias voltage created an unexpected difficulty: simply connecting a voltage source to each diode would place an AC ground in shunt with the variable capacitance. Concerns about speed of response aside, the circumstances would seem to call for a high-impedance voltage source. We pause when we consider the RC filter formed by the output impedance of the voltage source and the varactor's capacitance; this time constant must never be too large, lest the integrity of the control loop be compromised. The 1.8 kΩ resistor proved to be a satisfactory value.

---

6 The sole motivation for choosing a 250 MHz carrier was that the parts for such a frequency were readily available. At the time of this writing, it would have made far more sense to build a 1 GHz system. This was not possible without waiting 10 weeks or more for parts, however, and such a wait was incompatible with the length of the author's internship at Bell Laboratories.
Unfortunately these resistors, combined with parasitic elements in the DC blocking (27 pF) capacitor, inductor, and varactor diode itself, result in a termination for the quadrature hybrid coupler that is far from lossless. This non-ideality in the terminations has a direct impact on the loss of the entire phase shifter, as does the loss in the couplers themselves. The prototype exhibited attenuation as great as 4 dB for some phase values.

As alarming as this attenuation seemed, a more subtle concern is that this loss is not uniform across the range of capacitance. This is to be expected; at resonance of the series inductor-varactor circuit, for instance, the parasitic resistance in the varactor has a greater impact than in a non-resonance situation. The initial concern was that this would result in confusion for the controller because it introduces, in a sense, artificial fades that can only distort the system’s perception of the environment.

Neither of these two problems seriously disrupted the characterization of the prototype. The first of these problems (the very existence of attenuation, uniform or not) only constricted the range of power levels that could be achieved precisely by the controller. The second (the nonuniformity) was apparently not severe enough to be an issue. If it had been, we would have seen the controller become trapped in the local extrema of a new, multivariate transfer function:

\[ A = F(\phi(v_c),a(v_c)), \]  

(4-5)

where \( A \) is the amplitude of the combined signal, \( \phi \) is the phase relationship between the two antennas, \( a \) is the attenuation of the phase shifter, and \( v_c \) is the control voltage. We never observed this kind of entrapment.

There is one final concern for a system that uses this phase shifter: the controller should be intelligent enough to effect a phase “wrap around” when its output is at the borders of the control voltage range. One can imagine the controller becoming confused if, when its output to the phase shifter is at 4.95 volts, it determines that increasing the voltage still further will decrease the magnitude of the error. The proper course of action would be to jump to an equivalent phase (probably near a control voltage of 0 volts), and then continue normal operation from there. At present, the controller is not capable of handling this situation. The
characterization effort could still be carried out, however, subject to constraints on the size of phase adjustments that the system would be required to make.

**Design of the Power Detector**

As with the phase shifter, there was no interest in pushing the state of the art of power detector design. The chief concerns were appropriate speed of response and sensitivity. The heart of the circuit is shown in figure 4-4.

![Simplified diagram of power detector.](image)

**Figure 4-4:** Simplified diagram of power detector.

Strictly speaking, of course, this is the familiar “peak detector,” whose DC output is proportional to the square root of the peak input power. The distinction is more or less brushed aside in the literature; we will continue the practice in the following paragraphs.

The only real design problem associated with this circuit was making an appropriate choice for the RC time constant. It had to be long enough that an RF carrier (in our case, a 250 MHz signal) produced an approximately DC voltage, yet short enough that the fluctuations of the controller output, whereby it “felt” which direction was the proper one to go, were not muddied. But how to define “long enough” and “short enough”? It was decided that the former meant a decay of no more than 2% within the span of a period. The DC quality of the signal benefitted mightily from the bandlimit of subsequent amplifiers: built, as they were, of common op-amps, they were blind to any remaining ripple by a good two decades in frequency. This choice of RC time constant took care of the other concern too, as it was 250 times too small to meet the 2% criterion for a 1 MHz signal. This proved to be satisfactory.
Figure 4-5 is a schematic of the complete power detector.

It is seen that only minor modifications have been made to the circuit shown in figure 4-4: the 470 nH inductor, and the gain of ten amplifier. The placement of the inductor anticipates a blocking capacitor in the output stage of the RF source. If this is, in fact, the case, the inductor insures that there is always a DC path through which this capacitor can discharge. Theory predicts, and a poorly designed first prototype confirmed, that failure to provide this path results in a charge build-up on the coupling capacitor. The detector becomes unresponsive shortly after power is applied.

The amplifier following the detector is an LM310 buffer, trimmed for zero offset, and an LM301. One is tempted to worry that the input bias current of the buffer will badly affect the circuit, given the small size of the filter capacitor. A quick calculation provides much in the way of comfort: in 4 ns (the period of a 250 MHz carrier), the buffer draws a mere $2.7 \times 10^{-5}$ pC of charge, or 4.5 µV on a 6.2 pF capacitor. A more serious concern is getting approximately 1 MHz of bandwidth at a gain of 10. This was accomplished using the LM301. Taking advantage of the attenuation ladder in the feedback path, the compensation capacitor was reduced from its normal value by a factor of ten. This final value was experimentally determined.

Enough hardware had now been built to allow a meaningful test of the developments in Chapters 2 and 3. The details of our experiment are discussed in Chapter 5.
Chapter 4 References


Chapter 5: Experiment, and Results

Before contemplating the design of a meaningful simulation, one first wonders whether a device already exists for this purpose. Hewlett-Packard, at the time of this writing, has such an RF channel simulator available for purchase and for temporary loan. The user must provide two RF signals: the desired carrier, and a signal 6 MHz lower in frequency. The device uses this second signal as a local oscillator for a down conversion; the carrier is imbued with Rayleigh fading characteristics in baseband processing, and then an up conversion is performed to restore the carrier to its original frequency. We worked very hard to use this apparatus in our experiment, as it would prove beyond all doubt the ability of the controller to mitigate the effects of multipath propagation. It turned out not to be practical.

The problem was that the RF output contained artifacts of the conversion process that could not be easily filtered out. In addition to the faded 250 MHz signal that we desired, the output contained a strong 244 MHz signal (local oscillator feedthrough). Viewed on a spectrum analyzer, we saw that the strength of this feedthrough was a great as that of the desired carrier. This actually was not disastrous, as its power level did not fluctuate: its corruption of a broadband power measurement would thus be limited to the addition of an offset. We were less fortunate with a second artifact. The spectrum analyzer showed an “image” carrier at 238 MHz,
whose fading characteristics seemed *uncorrelated* with those of the main carrier. Again considering the broadband nature of the power measurement, one suspects this to be an insurmountable vice.

Of course, filtering was not a physical impossibility. There are design houses that specialize in designing filters of this type. Unfortunately, the turnaround time was too long given the time constraints of the internship.

We did try one other option, which was to use the spectrum analyzer as an extremely narrow-band power detector. It was possible to limit the display so that it only showed the power level at 250 MHz. It turns out that the Hewlett-Packard analyzers have a video output: a 0-1V analog signal that is proportional to the position of the trace on the screen. This failed to work too, however, as the new “power detector” lacked the ~1 MHz bandwidth necessary for the purposes of the controller. Our ingenuity exhausted, we turned to the task of designing an alternative, but hopefully as meaningful, experiment.

**The RF Experiment**

Figure 5-1 is a system diagram of the experiment that satisfied our requirement.

![Figure 5-1: RF experiment.](image)

- **HP8665B Synthesized Signal Generator**
  - Generates 250 MHz carrier that is AM modulated to imitate rapid fades.
- **Power Detector**
- **Analog Adaptive Controller**
- **Command Power Level**
This test setup represents the worst-case fading scenario (from a control standpoint), in that the separate antenna elements undergo completely correlated fading. By adjusting the command power level, a manually variable DC voltage source, we could evaluate the controller’s behavior.

It might be supposed that the use of two signal generators is preferable to the use of one; it would then be possible to simulate uncorrelated fading on each of the antenna channels, a more realistic situation. But there is strong, two-fold motivation for doing it as pictured in figure 5-1. First, as we have mentioned, a simultaneous fade on both antennas is the absolute worst case. It stands to reason that if the controller handles this properly, it will handle uncorrelated fading as well if not better. Second, in order for the experiment to work, the two function generators would have to be extraordinarily well matched. Consider the addition of two sinusoids that are close in frequency.

\[
A = \frac{1}{2} \cos \omega t + \frac{1}{2} \cos(\omega + \delta)t
\]

\[
= \frac{1}{4} (e^{j\omega} + e^{-j\omega}) + \frac{1}{4} (e^{j\delta} e^{j\omega} + e^{-j\delta} e^{-j\omega})
\]

\[
= \frac{1}{4} \left[ e^{j\omega} (1 + e^{j\delta}) + e^{-j\delta} (1 + e^{-j\delta}) \right]
\]

\[
= \frac{1}{4} \left[ e^{j\omega} e^{j\frac{1}{2}\delta} (2 \cos \frac{1}{2} \delta \omega) + e^{-j\omega} e^{-j\frac{1}{2}\delta} (2 \cos \frac{1}{2} \delta \omega) \right]
\]

\[
= (\cos \frac{1}{2} \delta \omega) \left[ \cos(\omega + \frac{1}{2} \delta) t \right]
\]

Finally, if \( \delta \) is small, we can approximate eq. 5-5 as

\[
= (\cos \frac{1}{2} \delta \omega) \left[ \cos \omega t \right].
\]

We see that even with no modulation of either carrier, we get an effective amplitude modulation at a frequency of \( \delta/2 \). This is alarming to contemplate, considering the bandwidth of the controller. We know that rejecting a 10 KHz disturbance is the best we can hope for: the two signal generators would have to be matched to within 80 parts per million. It may be possible to achieve matching that is better than this (e.g., it is sometimes possible to share a frequency reference), but it remains that the chosen method allows, in an important sense, a cleaner measurement.
From the standpoint of automatic gain control, the command power level can fall into one of three categories: greater than the available power, too low to be obtained (loss in the phase shifter renders complete cancellation impossible), and somewhere in between these two extremes. In the first and second case the controller should do its best, which is to adjust the received power to the appropriate extreme. In the third case, the variation of the power envelope with time should be flat. This is the behavior that we observed.

**Results and Discussion**

The results of our testing are summarized in table 5-1. The fading bandwidth of a car traveling at 60 MPH receiving a 2 GHz carrier is 360 Hz; knowing this helps to place the table in perspective.

<table>
<thead>
<tr>
<th>Type of modulation</th>
<th>Sinusoidal (Rejection bandwidth) KHz</th>
<th>Triangular (Indicates tracking speed) KHz</th>
<th>Square Wave (Recovery time) μsec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Fade (dB)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6.5</td>
<td>6.7</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>3.8</td>
<td>4.5</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>3.1</td>
<td>3.3</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>2.6</td>
<td>2.4</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>2.4</td>
<td>2.3</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 5-1: Results of fading experiment.

These data clearly show that our prototype is able to deal with rapid fading; to achieve a fading bandwidth of 2.4 KHz, the aforementioned car would have to travel at 400 MPH.

The interpretation of the data for sinusoidal modulation is relatively straightforward: it provides a clear indication of the controller’s ability to compensate for bandlimited fading. Notice that this bandwidth is lower than what we might have expected from the tests in Chapter 3. Those tests were done with the control voltage between 0 and 5 volts, as they are in this present case. But recall that it was the slope during a time step that determined tracking speed, and that here that slope varies significantly over the course of a modulation period.
The triangle wave data is provided for comparisons with similar tests in Chapter 3. The recovery time is also included for this reason, although it is useful in its own right in that shows how long one might be expected to wait after system power up to achieve the desired input signal level.

It instructive to look at the oscilloscope traces, by way of confirming that the controller is doing what it should. Curiously, the controller worked better with the fast transient detect circuitry disabled; in this experiment, as well as in the pictures to follow, only the slow adaptation system was left operational.

The first picture, figure 5-2, illustrates the system's behavior when the power command exceeds the total power available.

![Figure 5-2: Power command (top, flat line) set too high.](image)

We expect the system to maintain the highest received power possible; this value of the received power fluctuates sinusoidally, and it is clear that the controller is functioning properly. If we were to lower the power command a certain amount, we would enter a region in which the requested power is available some of the time; presumably this fraction of the time would increase as we lowered the command. Figure 5-3 illustrates this case.

![Figure 5-3: Power command achievable for a fraction of each modulation cycle.](image)

We get basically what we expect, but the size of the controller's vibrations in figure 5-3 are distressing. It is here that we see in graphic detail one of the perils of closing the loop around a
highly non-linear plant. Reexamination of equation 2-14 shows that when the signals are combined constructively, we have the added benefit of a small incremental gain in the phase-to-amplitude transfer function. This gives us the small step sizes, and therefore the clean waveform, that we see in figure 5-2.

Figures 5-4 through 5-7 summarize the performance of the system in various other situations.

**Figure 5-4:** Power command achievable throughout modulation cycle.

**Figure 5-5:** Phase control signal generated by controller for situation in Figure 5-4.

**Figure 5-6:** Power command too low.

**Figure 5-7:** Recovery from square wave modulation.

Figure 5-7 depicts conditions similar to those in figure 5-4, except with square wave, instead of sinusoidal, modulation. The “recovery time” measurements are, graphically, the width of these spikes at their bases.
Chapter 6: Conclusion

In closing, it is appropriate to review what is new about this work, as well as to propose ideas for its continuation. In so doing, an author implicitly suggests a unique place for his work in the sea of published literature. The hope is that the scientific community will decide it deserves to be there.

This project began as a research effort for commercial wireless products: the aim was to build a device to help hand-held and mobile receivers perform well in a Rayleigh fading environment. The author’s attention, however, was quickly arrested by a general class of feedback control problems, of which the present case was an example: designing a feedback system for a plant that rendered the sign of the loop transmission uncertain. Consideration of this problem led to the re-invention of a fairly old idea; Denidni and Delisle refer to it as the “on-off” algorithm in their 1995 paper, and a short reference trail leads to 1966 as the date of the earliest paper the author was able to find. It was a starting point, however, and provided some assurance that the idea was workable. The result of this early research was the basic, non-adaptive circuit discussed in Chapter 3. The circuit implementation is original: the early papers dealt with the algorithm itself, with physical systems as more or less an abstraction. The 1995 paper is an exception to this, but their approach differed fundamentally in that they used computer control.
The question remains, however: what is the value of the basic controller analysis in Chapter 3, given that the fundamental ideas can be found in earlier work? It is different in that it is the analysis of a real, physical system—we get to see how the controller deals with various non-idealities. The papers that have been cited dealt primarily with simulations. They have thus made a valuable theoretical contribution; the treatment of the basic controller in this paper necessarily leans toward the practical. Both perspectives are useful to an engineer.

Experiments on this early basic system were instructive, and prompted the contemplation of ways to beat tradeoffs that were inherent to the algorithm. It is here, in the ‘adaptive behavior’ sections, that Chapter 3 presents what we believe to be a genuinely new contribution.

With the completion of the RF prototype described in Chapter 4, we had a demonstration of the idea that started this whole project: a system that employed both constructive and destructive interference to perform the function of automatic gain control. Its performance was not perfect. Nevertheless, it served well as a proof-of-concept for the original proposed system (see figure 1-6). Current (two-element) diversity systems can require an LNA, a mixer, and a variable attenuator in each receiving branch: 6 expensive RF components. The proposed system would require only two, having eliminated the attenuators completely, and, by virtue of combining the signals in the RF stage, having eliminated one LNA and one mixer. This reduction is not without penalty, of course. We now need a directional coupler and a fast power detector with large dynamic range to make the system work. The controller itself could be redesigned as an IC, or have its function emulated by a process on a baseband processing chip: it is not expected to add significantly to the cost. We are convinced that the new receiver would be cheaper. It is too early to determine by how much.

**Further Work**

Most of the areas for additional work fall into one of two categories: improvement of the controller, and the development of the proposed system into a real system. Of the two, the controller work would be more appropriately classified as a “research” problem, as one imagines there are many untried optimizations, as well as alternate solutions, that have not been looked into. This is not to
trivialize the second category: there are no doubt a myriad of thorny implementation issues to be dealt with. Indeed, the time is past when, in the name of “proving a concept,” certain difficulties could be brushed aside. The issue now is whether our new system really is the cost-efficient alternative we believe it to be.

With regards to improving the controller, the main research issue now would have to do with taking adaptive behavior to the logical extreme. As has been mentioned, this would probably demand the incorporation of a microprocessor into system. The ideal would be an almost universal feedback controller that could adapt to any situation in which it found itself. In all likelihood this question has been raised before. Nevertheless, this basic algorithm offers the opportunity of making a fresh approach.

It is worth noting that one fundamentally new adaptive behavior would have to be added in order for this universal controller to be workable: it would have to adaptively determine the clock period. Chapter 2 alluded to this difficulty, when we mentioned that the algorithm demands extraordinarily high bandwidths of the plant in a system. We were fortunate in the present case, as the speed specifications for the complete loop were far slower than the bandwidth of the parts involved. This will not always be the case, and represents a huge barrier to the would-be designer of this system.

In considering the controller, though, we have gone somewhat afield of the main point which was to build a mobile receiver with superior fading performance at a lower cost. This project no doubt has a number of problems that cannot be anticipated at the outset, but merely dealt with as they arise. We know now that it is theoretically possible. A couple of problems that the author foresees:

- A fast, sensitive, power detector with a wide dynamic range will have to be built (or bought). This was one of the “certain difficulties that could be brushed aside” in the prototype, where input power was generally around 5 dBm. In a real system a strong signal will likely be at around -10 dBm after the LNA. An unobtrusive directional coupler will funnel only about 1/100th of this power to the detection branch.

- If the system is truly to be used in a hand held unit, a phase shifter that is small, low-loss, fast, cheap, and low voltage will have to be found. The phase shifter used in the prototype was fast. Its voltage
requirement of zero to 5 volts was not terrible, but the current push in the industry seems to be toward systems that use 3 volts or less.

It is anticipated that this work will be continued at Bell Laboratories in Murray Hill, NJ. The author hopes it will prove as useful to them as the last seven months have proven valuable, and memorable, to him.
Acknowledgments

I would like to express my gratitude to the following people, who at various times during the course of my brief academic career have been mentors to me in the truest sense of the word.

Dr. Jenshan Lin, Member of Technical Staff at Bell Laboratories, Lucent Technologies.

Dr. Lin was my advisor at Bell Labs for the duration of my stay there from June 1996 to January 1997. I would like to thank him for providing me with guidance when I needed it, and for trusting me enough to let me explore my own ideas. I wish him and his young family (congratulations on the new baby!) all the best.

Dr. James K. Roberge, Professor of Electrical Engineering at MIT

In addition to being one of his thesis advisees, I have taken four of his classes and served as a TA for one of them. I have nothing but admiration for his brilliant insight into the world of analog circuit design, and for his ability to teach. I consider my time spent learning from him to be the most formative academic experience of my career, and I would like to thank him for taking an interest in me. Thank you, Prof. Roberge.

Dr. Y. K. Chen, Department Head, High-Speed Electronics Research, Bell Laboratories, Lucent Technologies.

Among the many things he has done for me over the course of this project, I am most grateful for Dr. Chen’s support of a patent application on the final prototype. He risked considerable embarrassment when he took me around to various development engineers to present my ideas; that took a great deal of faith in my abilities. I would like to thank him for giving me such an astounding opportunity, and for a wonderful time spent doing research in his department.

Dr. Cherry Murray, Department Head, Semiconductor Physics Research, Bell Laboratories, Lucent Technologies.

Dr. Murray is the person to whom I think I owe the most. I do not understand why she decided to look after me that summer following my freshman year in college. Since then, she has treated me with only kindness, generosity, and in the spirit of true mentorship. She has my undying gratitude, and my promise to make the most of the numerous opportunities she has given me.