MINIMUM-DELAY HF COMMUNICATIONS

BY

TOROS ARIKAN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Professor Andrew C. Singer

ABSTRACT

HF radio communications offer latency advantages over fiber-optics and are more economical than satellite links in long-distance communication applications. As a result, there is renewed interest in the field of HF from the high-frequency trading industry, as a means of transmitting financial data. Ideally, decisions should be made on received symbols as soon as they arrive, without waiting for other samples to accumulate. A significant problem is that the ionospheric HF channel is a dynamic medium that is considered one of the most challenging communication channels. In this thesis, we investigate alternative techniques towards reliable communication with minimum delay. We study HF channel models and communication link design tools, which indicate that our minimum-delay problem is feasible but difficult to solve. Extant HF modems are observed to add prohibitive amounts of latency to ensure robustness, motivating us to design receivers that provide good performance given a delay constraint. We adapt a coupled MAP-RLS receiver studied in literature to the HF channel to yield minimal decision-making delay. We introduce a multitrellis adaptive Viterbi algorithm (MAVA) that solves the problem of equalization for sparse and time-varying ISI channels, producing a robust MLSE receiver with several milliseconds of decision-making delay. Finally, we cascade this MLSE receiver with MAP detection to obtain high-fidelity low-delay performance, finding a practical solution to minimum-delay HF communications.

Dedicated to my grandfather, Omer Fikri Arikan.

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Andrew C. Singer, for all his hard work and generosity. I will always remember his lessons in dedication, professionalism, and perfectionism.

I would also like to acknowledge Jump Trading, LLC, for providing financial support, an incredible internship experience, and above all, a great research opportunity.

Finally, I would like to express my gratitude towards my friends and family, who continue to amaze me with their boundless goodness and kindness.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

High frequency (HF) radio communications are a well-established field within electrical engineering, with applications in radio broadcasting, naval communications, and secure military communications [1], [2]. Within the 3-30 MHz HF frequency band, it is possible to reflect a radio transmission off the Earth's ionosphere. Such skywave propagation can have a range of thousands of kilometers with multiple hops, or reflections, between the ionosphere and the Earth's surface.

There is renewed interest in HF communications from high-frequency trading firms, who wish to communicate financial data across the Atlantic ocean with as little latency as possible so as to gain a competitive edge. Currently, transatlantic data communications are accomplished with relatively slow fiber-optic cables, or with expensive satellite links. HF data links would deliver the free-space speed-of-light propagation advantages of satellite links over that of fiber-optics, at a fraction of the cost. The disadvantage is working with a difficult doubly-dispersive channel, subject to seasonal and daily cycles, fading, and occasional sudden ionospheric disturbances due to solar flares.

Optimization over delay is critical for high-frequency trading applications: if a decision is made at the receiver only milliseconds after that of a competitor, then the transmission has failed its purpose, regardless of system reliability or data rate. Fast transatlantic fiber-optic cables have between 30 and 60 ms of latency, while HF communications have about 20 ms of latency on the same path, thus yielding a tight delay budget. Although

there are many critical design steps in a practical HF communications link, including site selection, antenna selection, and statistical channel analysis, we observe that the latency bottleneck due to decision-making delays at the receiver poses the greatest challenge. Well-established narrowband HF communication systems incorporate hundreds of milliseconds of interleaving and deinterleaving, with comparable delays due to forward error correction. The resulting delays render current HF data links slower than fiber-optics. Hence, we need to design new minimum-delay receivers that are as reliable as possible given our latency constraints.

In this thesis, we motivate, derive, and simulate HF receivers that are optimized for minimum latency. Our major contributions are a multitrellis adaptive Viterbi algorithm (MAVA) that provides robust maximum-likelihood sequence estimation (MLSE) capability on HF channels with several milliseconds of latency; and a hybrid MLSE-MAP receiver that allows us to make decisions on received symbols without additional delay, and with near-optimal performance. Both of these methods involve simultaneous estimation of transmitted symbols and channel model coefficients. The novelty of our design constraint allows us to combine and extend well-established algorithms in novel ways.

The thesis is structured as follows. In Chapter 1, we provide background research on HF link design, channel modeling, and HF receiver designs. In Chapter 2, we present our receivers, highlighting the considerations that drive each step of our design process. In Chapter 3, we present simulation results for each receiver, comparing and contrasting their performance to decision-making on a perfectly known channel. In Chapter 4, we present a conclusion and propose future avenues for research.

1.2 Channel Considerations for Minimum-Delay HF Communications

The physical properties of the Earth's ionosphere allow long-distance sky-wave communications to be conducted through reflections of HF transmissions; and the state of the ionosphere at a given time and place is the most

important factor in determining reception quality. For this reason, the dynamic behavior of the ionosphere has been extensively researched since the inception of wireless communications [3]. We currently have statistical and computational tools that can evaluate the long-term performance of a hypothetical HF link, allowing us to evaluate the feasibility of our designs. We also have tapped delay line models for the instantaneous HF channel profile, which is essential for proper receiver design. In this section, we review these models as they pertain to our minimum-delay HF problem; we consult [4] and [5] for a thorough treatment of this subject.

1.2.1 Feasibility Analysis for Minimum-Delay Transatlantic HF Communications

The design of an HF communication link is dictated by the setting and difficulty of the application. Intermittent point-to-point speech broadcasting can be achieved with a small ham radio and antennas that are several meters long, while reliable transatlantic HF radio broadcasting requires large antenna arrays more than 100 meters long, and power on the order of hundreds of kilowatts [6]. The practical application that we are designing for is continuous transatlantic HF data transmission throughout the day, with data rates on the order of tens of kbps, and a bit-error rate (BER) on the order of 10^{-3}. To the best of our knowledge, the implementation of such an HF link has not been reported due to the inherent difficulties of long-range HF data transmission. Therefore, we must first determine the feasibility of minimum-delay HF communications, and what it would take to accomplish this goal.

A useful and well-recognized program for predicting HF link performance is VOACAP [7], which uses archived performance data to produce statistical estimates of signal-to-noise ratio (SNR), coverage, and the best operating frequencies throughout the day. Monthly median values are used to produce these estimates; most significantly for us, the SNR is defined as the monthly median signal power (dBW) picked up at the receiver minus the monthly median noise power (dBW), for a given HF link. The program calculates the SNR that could be achieved if the entire signal power were transmitted

within a bandwidth of 1 Hz, thus limiting the noise power to that observed within 1 Hz. The user is expected to scale this SNR value to their own desired bandwidth by assuming white noise, so that the noise power in a 10 kHz bandwidth is greater than the noise power within 1 Hz by a factor of 10^4. Thus, if we wished to transmit at a bandwidth of 10 kHz with a given signal power, and VOACAP provided us with an SNR estimate of 60 dBW, we should expect to observe 20 dBW of SNR at 10 kHz [1]. Due to the use of monthly statistics and sweeping assumptions in estimating SNR, VOACAP's outputs should be used for long-term link design, but not as a measure of instantaneous link quality.

The inputs to VOACAP that concern us in this preliminary analysis are the monthly smoothed sunspot number (SSN), the transmitter and receiver locations, the noise floor, the required circuit reliablility, the power supplied to our transmitter, and our choice of transmit and receive antennas. We perform two simulations with the following parameters to determine our requirements for a long-term minimum-delay link:

- Transmitter/receiver locations: Champaign, IL (40N, 88W) to London, UK (51.5N, 0.17W). These locations are selected for illustration purposes; actual locations should be in rural areas.
- Noise floor: -150 dBw/Hz, corresponding to a typical rural area.
- Required SNR: 60 dB. This corresponds to 20 dB of SNR at 10 kHz, under the previous assumptions.
- Required circuit reliability: 80%. This is the estimate of the percentage of days in a given month when the signal quality is acceptable. It is difficult to maintain an HF skywave link unbroken for an entire month; there will be days of disturbed channel conditions when the data link is intermittent.
- Transmit antenna: HR 4/4/1, a multiband aperiodic reflector array. This is a sample large transmit array; a specific high-directivity (> 18 dBi gain) antenna array should be used.
- Transmit power: 50 kW.

- Receive antenna: Isotropic, 10 dBi gain. This theoretical model antenna stands for a highly directive (8-10 dBi) receiver array, to be determined in the design process.

We run simulations for January 2011 (SSN: 46), a relatively calm month, and December 2011 (SSN: 93), a relatively disturbed month that nevertheless has on average a more stable channel with a more reflective ionosphere. These simulations, presented in Figures 1.1 and 1.2 respectively, illustrate the SNR that we can expect to achieve with the link. Monthly estimates of median SNR are calculated across the range of HF frequencies throughout a given day, and the results are expressed as a contour plot. We observe that we can reliably attain 20 dB SNR at 10 kHz throughout the day for January, and can reliably attain more than 30 dB SNR throughout December. These simulations indicate that a well-designed HF link with highly directive transmitter and receiver arrays and 50 kW power can expect to attain 15-35 dB SNR throughout the day for most days in a given month. We will later observe that minimum-delay decision-making is feasible within this SNR range.

Another key consideration in our link analysis is the transmission time, which has to beat the latency of fiber-optics by a comfortable margin. Ionospheric conditions, the takeoff angle of the transmitter beam, and the transmission frequency all come together in determining the number of hops the transmission will take, and are therefore critical parameters in determining latency. However, current HF simulators do not simultaneously conduct electromagnetic wave propagation analysis and channel performance predictions. Typically, such simulation packages use the hourly and monthly median trends in their analysis, and do not take into account the state of the ionosphere at a given time. IONOLAB-RAY [8], [9], [10] is a recently developed ray propagation algorithm that bridges this gap, by modeling electromagnetic wave propagation as a ray that travels in the ionosphere, partitioned into 3-D voxels. The algorithm also calculates channel parameters such as attenuation, time delay, group velocity, phase velocity and the Faraday rotation of the electromagnetic wave. In order to estimate the time delay of an HF link between the US and Europe, we run a set of simulations using IONOLAB-RAY where the transmission frequency, azimuth and elevation

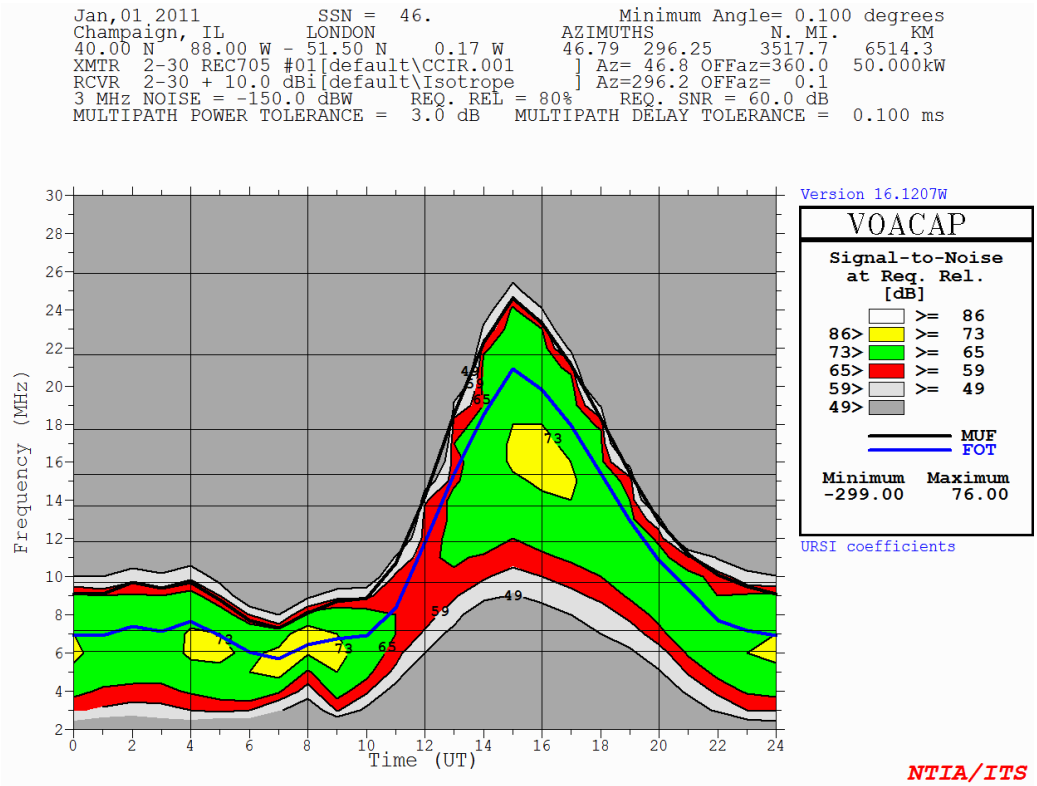

Figure 1.1: VOACAP simulation for SNR in dB-Hz, throughout a given day in January 2011.

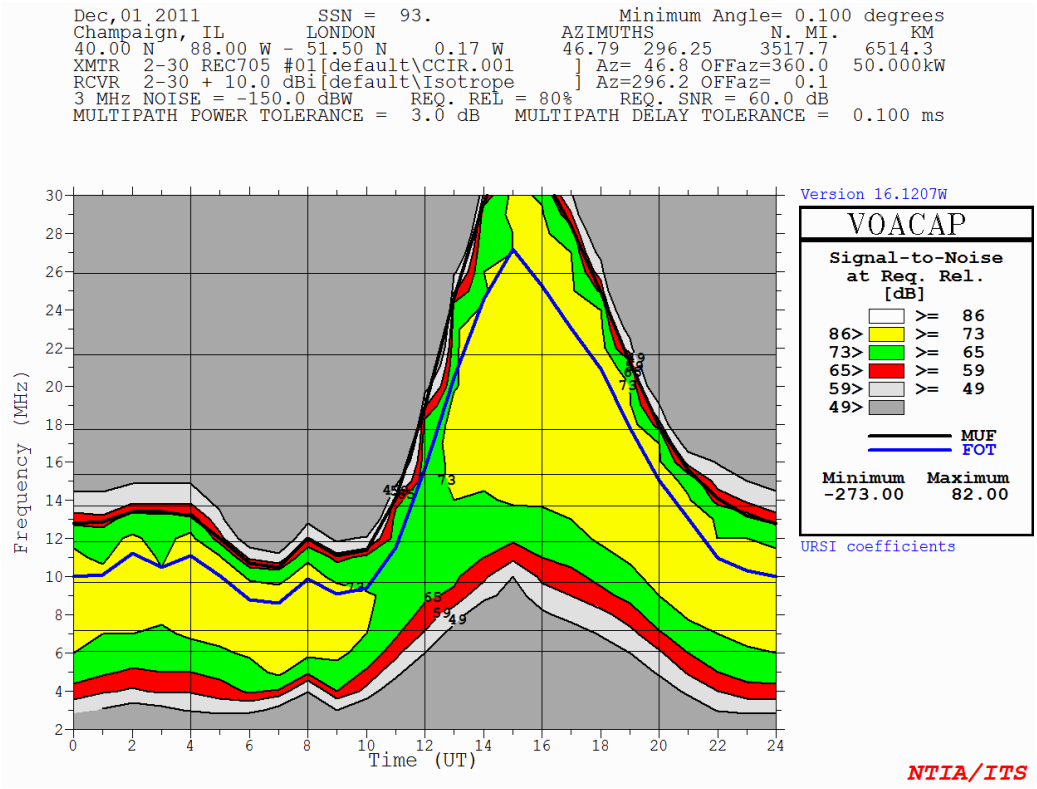

Figure 1.2: VOACAP simulation for SNR in dB-Hz, throughout a given day in December 2011.

of the electromagnetic wave exiting the transmitter, the time of transmission, and the state of the ionosphere are our modified parameters. We run IONOLAB-RAY for the days of 14 and 17 April 2011 (nondisturbed days) and 24 and 25 April 2011 (disturbed days) for the hours of 00:00 UT to 22:00 UT, using transmission frequencies between 10 and 25 MHz, and elevation angles from the local horizon of the transmitter between 0.1 and 60.5 degrees. We observe that London can be reached from Champaign in 3 hops, with a time delay of 20 to 25 ms. Older transatlantic fiber-optic connections have latencies between 59 and 65 ms [11], while new Chicago to New York and New York to London connections that serve high-frequency trading companies feature 7 and 30 ms of latency respectively [12]. As a result, the HF link can count on a delay advantage of more than 10 milliseconds - a wide margin from a high-frequency trading perspective that makes our project a feasible and worthwhile endeavor, but also narrow enough from a communications perspective to pose a novel research challenge.

1.2.2 Channel Model

The size and cost of HF antennas and broadcast licensing issues make it difficult to carry out HF experiments in the field, especially for transatlantic communications. As a result, a great deal of effort has gone into generating HF channel models for standardized testing of modems [13], [14], [15], [16], [17]. In this section, we briefly cover the Watterson model [13], and establish the discrete-time baseband notation which we will use in our derivations. We refer to [18] for a thorough overview of the theory and implementation of HF channel models.

The Earth's ionosphere is composed of ionized layers in the atmosphere, and is a time-varying medium that follows seasonal and daily cycles. An HF beam can be reflected off these layers to achieve over-the-horizon radio communication, potentially over thousands of kilometers. The HF channel is a time- and space-varying, spatio-temporally dispersive, multipath, frequency selective fading channel where the propagating electromagnetic waves are subjected to attenuation, absorption, refraction (time delay), frequency shifts (Doppler) and polarization shifts (Faraday rotation). The different propa-

gation modes result in multipath components separated by milliseconds of delay. The dynamic movement of ionospheric layers results in the Doppler spread of the received signal; and since the reflection of the HF beam occurs off of many randomly moving ions, this Doppler spread has a Gaussian distribution. The Watterson model approximates the narrowband HF channel by a tapped delay line, where each multipath component is independent and Rayleigh fading, with a Gaussian Doppler spread, and subject to additive white Gaussian noise (AWGN) [18]. We express these properties in our notation as follows.

For an m-path channel where the i^{th} path exhibits the time-varying complex gain $c_i(t)$ at delay τ_i, the received signal $r(t)$ in baseband is given by:

$$r(t) = \sum_{i=0}^{m-1} c_i(t)\, s(t-\tau_i) + N(t), \tag{1.1}$$

where $s(t-\tau_i)$ is the value of the transmitted signal at time $(t-\tau_i)$, and $N(t)$ is complex additive white Gaussian noise with variance σ^2. We assume that the channel is quasi-wide sense stationary, so that the second-order statistics of channel taps $c_i(t)$ are assumed constant over a reasonable time interval during which communications take place. We also assume that the multipath components are independent, which is valid if the bandwidth is small enough that correlated scatterers with different delays cannot be resolved at the receiver [19]. With these assumptions, the autocorrelation of each tap is defined as:

$$R_c(u, \tau_i) \triangleq \mathrm{E}\left[c_i(t)\, c_i^*(t-u)\right] \tag{1.2}$$

The Watterson model assumes a Gaussian distribution for the Doppler spread. For a Doppler spread variance of σ_ν^2, the Doppler spectrum of a tap is given by:

$$S_h(\nu) = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} e^{-\frac{\nu^2}{2\sigma_\nu^2}} \tag{1.3}$$

The Doppler spectrum is defined as the Fourier transform of the autocorrelation of each tap gain. The autocorrelation function that corresponds to the Doppler spectrum in Equation (1.3) is:

$$R_h(u) = e^{-\frac{(2\pi\sigma_\nu u)^2}{2}} \tag{1.4}$$

A wide Doppler spectrum corresponds to a narrow autocorrelation function. This indicates that close samples are less correlated, and that the channel is therefore varying faster in time [20]. When the ionosphere is in a disturbed state, the Doppler spread is greater and the channel variation is faster.

The Watterson model assumes that the narrowband HF channel exhibits Rayleigh fading, meaning that we model our channel taps $c_i(t)$ as complex zero-mean Gaussian processes, so that the envelope $|c_i(t)|$ is Rayleigh distributed. We derive the equivalence of these two properties as follows.

We observe that for a single tap:

$$c(t) = r(t)\, e^{j\phi(t)} = x(t) + jy(t), \tag{1.5}$$

where $x(t)$ and $y(t)$ are jointly Gaussian random processes that are independent, zero-mean, and distributed with the same variance. Then, their joint probability density is given by:

$$P(x,y) = P(x)\, P(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1.6}$$

We transform differential areas with the change of variables $dx\, dy = r\, dr\, d\phi$, to obtain the joint pdf of r and ϕ as:

$$P(r,\phi)\, r\, dr\, d\phi = P(x)\, P(y)\, dx\, dy = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\, r\, dr\, d\phi \tag{1.7}$$

Then by integrating, we obtain:

$$P(r) = \int_{-\pi}^{\pi} \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\, d\phi = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \tag{1.8}$$

$$P(\phi) = \int_{0}^{\infty} \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\, dr = \frac{1}{2\pi} \tag{1.9}$$

Hence, the magnitude of each channel tap has the Rayleigh distribution,

and the phase has the uniform distribution.

For standardized testing of HF modems, the International Telecommunications Union radio communications group (ITU-R) recommends simulating a two-tap HF channel according to the Watterson model, with tap gains $c_0(t)$ and $c_1(t)$ that are fixed in location and separated by delay d, and subject to independent Rayleigh fading with identical Doppler spreads of $2\sigma_\nu$ [21]. The tap separation and Doppler spread parameters are assigned based on approximate geographical location and channel conditions. Since the problem we approach is transmission of financial data across the Atlantic, we perform our simulations with midlatitude conditions, given in Table 1.1.

While the Watterson model is a useful tool for the analysis of our receiver designs and detection algorithms, it has a number of shortcomings that must be addressed by a practical HF modem. Most importantly, the Watterson model assumes that the received symbol is corrupted by simple AWGN, but the HF channel also suffers from impulsive noise due to ionospheric disturbances and occasional bursts of interference. Bursts of noise up to 1 second long can severely disrupt data links. As a result, experimental results from wideband soundings of the channel must also be used for modem design. Furthermore, in the absence of long forward error correction, the only way of mitigating these long bursts would be to leverage channel and frequency diversity in the form of a second antenna transmitting the same symbols from a different location and at a different frequency. Another key shortcoming of the Watterson model is that it is only considered valid for bandwidths smaller than 12 kHz. Beyond this bandwidth, the Watterson model may or may not be valid, depending on channel conditions. We conduct our simulations at a symbol rate of 10 kbps so that the Watterson model can be applied without complications.

Table 1.1: ITU-R channel parameter recommendations for simulations.

Midlatitude Conditions	**Delay Spread,** d	**Doppler Spread,** $2\sigma_\nu$
Calm	0.5 ms	0.1 Hz
Moderate	1 ms	0.5 Hz
Disturbed	2 ms	1 Hz

Having assumed the Watterson model for our channel, we are able to establish the operation of our communication link in baseband. At time k, the first arrival from the channel has a complex coefficient c_{0k}, and the second arrival is delayed by d and has a coefficient c_{1k}. The received symbols are corrupted by samples of circularly-symmetric complex Gaussian noise N_k, of variance σ^2. The transmitted symbols s_n are produced by a simple modulation scheme; for good performance and simplicity in decision-making, we generally use binary phase-shift keying (BPSK) modulation, so that $s_n \in \{\pm 1\}$. As a result, the received symbol r_k is given by:

$$r_k = c_{0k}s_k + c_{1k}s_{k-d} + N_k \tag{1.10}$$

In order to make an accurate decision on s_k without delay, we need information on s_{k-d} as well. This information is provided by r_{k-d}; taking the channel variations into account we have:

$$r_{k-d} = c_{0,k-d}s_{k-d} + c_{1,k-d}s_{k-2d} + N_{k-d} \tag{1.11}$$

Equations (1.10) and (1.11) justify our strategy for minimum-delay decision-making. If we have both the accurate estimates for the channel coefficients and the hard decisions on past symbols s_{k-d} and s_{k-2d}, we can find a reliable estimate for $s_k \in \{\pm 1\}$ in the presence of Gaussian noise.

1.3 Traditional HF Modem Design

An essential step in approaching the minimum-delay HF communications problem is to review the standard HF communications techniques used in literature, and to observe how they tackle the problems of multipath and fading. HF has been used for data transmission for over 60 years, especially for secure military communications, so that we have a number of proven protocols to study [22], [23], [24]. Motivated by the critical importance of reliability to such applications, traditional HF modem design emphasizes robustness and features long processing delays.

High data rate (3-10 kbits/s) HF modems employ decision feedback equalizers (DFE) to mitigate intersymbol interference (ISI) from the received sig-

nal [18], and transmit frames of training symbols followed by data. A DFE structure is necessary due to the deep spectral nulls induced by the fading channel. Setting the reference tap of the equalizer to T_r, in samples, the total delay of the equalizer in symbols is given by:

$$1 + \frac{T_r}{(\#\text{ of samples per symbol})} \tag{1.12}$$

Generally, a recursive least-squares (RLS) equalizer is employed to track channel variations, whose fast convergence is needed to combat rapid fading. However, a basic receiver of this form can be sensitive to noise, so unless we have an accurate channel estimate on which we can directly calculate equalizer coefficients, high SNR will be necessary for acceptable performance [25]. Incorrect decision feedback can cause this RLS DFE equalizer to diverge, requiring periodic restarts and generally resulting in poor performance [26]. Therefore, interleaving/deinterleaving of hundreds of milliseconds is often employed to mitigate bursts of interference or noise in the channel [18]. Forward error correcting codes and more sophisticated techniques such as turbo equalization [27], [28] are used to correct errors in the decoded sequence. Even though the data rate is high, large latency is needed for decoding purposes.

Low data rate (100 bits/s) modems can employ RAKE receivers, a technique originally developed for HF communications [29], to obtain the channel profile and coefficients. Little ISI results at these low rates, so we can delay and sum the multipath arrivals to improve SNR. This process also boosts the signal energy as we add multipath components rather than subtract their effects. In practice, the data bits are transmitted alongside a pseudonoise (PN) sequence with a much higher symbol rate (10 kHz-1 MHz). The received signal is correlated with delayed copies of the PN sequence, with integration times of hundreds of milliseconds, to yield the multipath arrivals and the corresponding channel coefficients. This technique is extremely effective at dealing with low SNR and disturbed channel conditions [30], at the cost of low data rate and high-latency communications.

Our goal is to make decisions on the received symbols either as soon as they arrive, or with only milliseconds of delay. Since we also wish to commu-

nicate with data rates on the order of 10 kbps, we are faced with a difficult communications problem requiring use of all available knowledge for decoding purposes. Knowing the channel profile, we would be able to determine which past symbols cause ISI; and if we also know the corresponding channel coefficients, we would be able to estimate the transmitted symbol from the received samples. Rather than obtain the channel coefficients through correlation, with long integration times, we can employ tracking based on training symbols and hard decisions. These considerations motivate maximum a posteriori (MAP) or maximum likelihood (ML) decision-making approaches that make use of past decisions to mitigate ISI, and thus avoid long delays. We will observe that by combining ISI mitigation and real-time channel tracking in this manner, we can accurately make minimum-delay decisions.

CHAPTER 2

MINIMUM-DELAY RECEIVER DESIGNS

2.1 Overview

In this chapter, we design three receivers that can be used in a minimum-delay HF modem. We justify these designs from a practical engineering standpoint, in addition to theoretical derivations well-established in literature. We propose a novel maximum-likelihood sequence estimation (MLSE) algorithm called the multitrellis adaptive Viterbi algorithm (MAVA), which allows us to apply MLSE methods to channels where wide tap separations would prevent the use of conventional MLSE algorithms. We also present a novel receiver design where we cascade an MLSE estimation stage and a MAP detection stage to produce low-delay decisions with strong performance.

An interesting feature of the minimum-delay HF communication problem is that we can easily run a fiber-optic cable between the transmitter and receiver, for delayed but practically perfect communication between the two. This fiber-optic link would have tens of milliseconds of additional latency over the HF channel, so we cannot make direct use of it in our receiver designs due to minimum-delay constraints. However, this feature would be invaluable in the field for a practical HF modem. It would allow us to accurately monitor channel conditions and modem performance, and should be incorporated into any protocol designed for high-frequency trading applications over the HF skywave channel.

The receiver designs proposed in this chapter feature sample-based processing instead of frame-based processing. There are several reasons for this choice. First, the large delay spread of the HF skywave channel means that accumulating a large frame of symbols results in too much latency. Con-

sider the ITU-R disturbed test channel, which has two taps that are $d = 2$ ms apart in delay. In order to accumulate six symbols s_k, s_{k+d}, ... , s_{k+5d} that are related to each other at the receiver due to multipath, we would need to wait for $5d = 10$ ms before processing, which is an unacceptable latency for our target application. Second, there is no proven protocol yet for robust minimum-delay HF skywave communications. The optimal lengths of the training and data frames for different channel conditions have yet to be experimentally determined; furthermore, while the data symbols will be transmitted frame by frame in a practical software-defined radio implementation, the exact lengths of the frames and the processes by which they will be generated will vary between applications. As a result, we present receiver designs that are applicable for all possible modem implementations. We note that frame-based processing is significantly less complex than sample-based processing for many receiver designs, especially for those that incorporate the Viterbi algorithm. As a result, all of the receivers we design here are directly applicable to frame-based processing, and only require simplifications to what we propose.

2.2 Coupled MAP-RLS Receiver

To tackle the problem of minimum-delay decision-making, we first focus on designing a receiver that can reliably make decisions on the transmitted symbol s_k at time k, thus yielding zero delay due to the decision-making process. This would hold great importance for HFT applications. In this section, we propose a receiver that couples zero-delay MAP detection and RLS channel tracking for BPSK signaling over the two-tap Watterson channels of the ITU-R model. The extension of the receiver to more complex channel structures and higher-order constellations is straightforward. We observe that this receiver matches extant designs in literature, and serves as our baseline case for performance comparisons.

The maximum tolerable delay requirement in our HF receiver imposes a severe restriction on error-correcting capabilities and noise tolerance. For acceptable performance, we would like to leverage all the information we have pertaining to the received signal. We therefore cast our decision-making

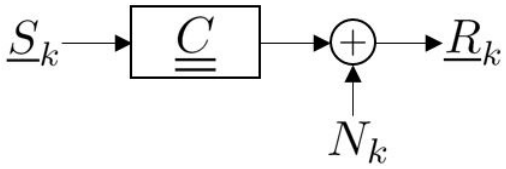

Figure 2.1: End-to-end communication model used to formulate the MAP detector.

problem in a coherent detection context, where in addition to our knowledge of the channel, we can exploit both hard decisions on previous symbols, and soft information in the form of the bit-error rate (BER) of the communication system. This reduces the crux of our receiver design problem to the formulation of an appropriate MAP detector.

We make the baseband approximation that the received signal is generated by the linear system given in Figure 2.1. Assuming the two-tap Watterson channel model, the received signal $\underline{r}_k$, the channel coefficients $\underline{\underline{C}}$, and the transmitted BPSK symbols $\underline{s}_k$ can then be expressed as:

$$\underline{r}_k = \begin{bmatrix} r_k \\ r_{k-d} \end{bmatrix}, \underline{\underline{C}} = \begin{bmatrix} c_{0,k} & c_{1,k} & 0 \\ 0 & c_{0,k-d} & c_{1,k-d} \end{bmatrix}, \underline{s}_k = \begin{bmatrix} s_k \\ s_{k-d} \\ s_{k-2d} \end{bmatrix} \tag{2.1}$$

The transmitted information is corrupted by additive circularly symmetric Gaussian noise with variance σ^2:

$$\underline{r}_k = \underline{\underline{C}}\,\underline{s}_k + \underline{N}_k, \tag{2.2}$$

$$\underline{N}_k = \underline{r}_k - \underline{\underline{C}}\,\underline{s}_k \sim \mathcal{CN}\left(\underline{0},\, \sigma^2 \underline{\underline{\mathrm{I}}}\right) \tag{2.3}$$

We assume that each transmitted BPSK symbol is equally likely. However, for past symbols s_{k-d} and s_{k-2d}, we possess hard decisions $\hat{s}_{k-d}$ and $\hat{s}_{k-2d}$. We also know that these past decisions were faulty with probability ϵ, which is approximated by the BER of the communication system at given channel conditions, and is assumed to be known in practice, using the delayed fiber-optic transmissions of the true symbols. We thus have a prior on $\underline{s}_k$ as:

$$s_k = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases},$$

$$s_{k-d} = \begin{cases} \hat{s}_{k-d} & \text{w.p. } 1-\epsilon \\ -\hat{s}_{k-d} & \text{w.p. } \epsilon \end{cases},$$

$$s_{k-2d} = \begin{cases} \hat{s}_{k-2d} & \text{w.p. } 1-\epsilon \\ -\hat{s}_{k-2d} & \text{w.p. } \epsilon \end{cases} \tag{2.4}$$

The maximum a posteriori (MAP) estimate for $\underline{s}_k$ given $\underline{R}_k = \underline{r}_k$ is then obtained as:

$$\underline{s}_{k,\text{MAP}} = \underset{\underline{s}_k}{\text{argmax}}\, P_{\underline{S}_k|\underline{R}_k}\left(\underline{s}_k \mid \underline{r}_k\right), \tag{2.5}$$

$$P_{\underline{S}_k|\underline{R}_k}\left(\underline{s}_k \mid \underline{r}_k\right) = \frac{P_{\underline{R}_k|\underline{S}_k}\left(\underline{r}_k \mid \underline{s}_k\right) P_{\underline{S}_k}\left(\underline{s}_k\right)}{P_{\underline{R}_k}\left(\underline{r}_k\right)} \tag{2.6}$$

$$\Rightarrow \underline{s}_{k,\text{MAP}} = \underset{\underline{s}_k}{\text{argmax}}\, P_{\underline{R}_k|\underline{S}_k}\left(\underline{r}_k \mid \underline{s}_k\right) P_{\underline{S}_k}\left(\underline{s}_k\right) \tag{2.7}$$

where we have used Bayes' rule in Equation (2.6) and the independence of $P_{\underline{R}_k}\left(\underline{r}_k\right)$ from $\underline{s}_k$ in Equation (2.7). Due to our assumption of complex Gaussian noise, we know that $P_{\underline{R}_k|\underline{S}_k}\left(\underline{r}_k \mid \underline{s}_k\right)$ is given by:

$$P_{\underline{R}_k|\underline{S}_k}\left(\underline{r}_k \mid \underline{s}_k\right) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{\left\|\underline{r}_k - \underline{\underline{C}}\,\underline{s}_k\right\|_2^2}{\sigma^2}\right\} \tag{2.8}$$

We define a function $f\left(u, v\right)$ as:

$$f\left(u, v\right) = \begin{cases} 1-\epsilon & \text{if } u = v \\ \epsilon & \text{if } u \neq v \end{cases} \tag{2.9}$$

Then, using our knowledge of priors for transmitted symbols, we express $P_{\underline{S}_k}\left(\underline{s}_k\right)$ as:

$$P_{\underline{S}_k}\left(\underline{s}_k\right) = \frac{1}{2} f\left(s_{k-d}, \hat{s}_{k-d}\right) f\left(s_{k-2d}, \hat{s}_{k-2d}\right) \tag{2.10}$$

Therefore, by ignoring the constants $1/2$ and $1/2\pi\sigma^2$, we obtain $\underline{s}_{k,\mathrm{MAP}}$ as:

$$\underline{s}_{k,\mathrm{MAP}} = \underset{\underline{s}_k}{\operatorname{argmax}} \exp\left\{-\frac{\left\|\underline{r}_k - \underline{\underline{C}}\,\underline{s}_k\right\|_2^2}{\sigma^2}\right\} \cdot f\left(s_{k-d}, \hat{s}_{k-d}\right) f\left(s_{k-2d}, \hat{s}_{k-2d}\right) \tag{2.11}$$

For BPSK modulation, the symbol estimate is given by:

$$\underline{s}_{k,\mathrm{MAP}} = \underset{s_k, s_{k-d}, s_{k-2d} \in \{\pm 1\}^3}{\operatorname{argmax}} \exp\left\{-\frac{\left\|\underline{r}_k - \underline{\underline{C}}\,\underline{s}_k\right\|_2^2}{\sigma^2}\right\} \cdot f\left(s_{k-d}, \hat{s}_{k-d}\right) f\left(s_{k-2d}, \hat{s}_{k-2d}\right) \tag{2.12}$$

The extension of the MAP approach to more than two significant arrivals is straightforward, and involves the introduction of additional priors, the appropriate resizing of the signal vectors and channel matrix, and the maximization over more than 3 symbols.

One of the important parameters that must be accurately determined for correct operation of this receiver is the value of σ^2, the variance of the Gaussian noise added by the channel. In our simulations, we assume perfect knowledge of σ^2. In practice, we would need to dynamically determine the SNR; a relatively easy task during training frames, and a harder one during data transmission. There are many techniques in the estimation literature for calculating the variance of Gaussian noise [31]. A robust technique of noise estimation should be selected in the field after extensive experimentation, over different channel conditions, in order to quantify their practical limits and to determine how rapidly the noise floor changes.

A second key assumption for this receiver is accurate knowledge of the tap locations in the multipath channel. Even under the best channel conditions, we would expect the tap locations to vary, or drift slightly over time. New taps would also emerge over the course of receiver operation. The channel profile can be precisely determined by a sounding, and accurately tracked during our transmission, using well-established techniques in literature [32] [33]. As with the case of noise estimation, we would need to experiment with the actual channel over a period of time, trying out different techniques to figure out which provide the best performance for the real channel.

The derivation above also assumes that we have knowledge of the error probability ϵ, which would be difficult to estimate in a real-time system with a rapidly time-varying channel. Fortunately, we observe in simulations that the MAP decision is highly insensitive to our assumed or calculated value of ϵ, so that the receiver appears to be robust in this regard. Taking ϵ on the order of 0.01 is sufficient for our purposes; any large changes in this value are observed to yield severe increases in BER. The critical quantities for the MAP approach are the channel coefficients $\underline{\underline{C}}$, which have to be accurately known and updated.

In the next chapter, our simulations confirm a known feature of HF sky-wave communications: that delayed channel estimates, however accurate, are not serviceable for practical HF communications for moderate or disturbed channel conditions. To obtain the best performance possible, we have to update our channel estimate in real-time, using the current and past decisions. The resulting receiver has the hard decisions of the MAP decision-making module fed into a channel estimator that employs a reliable tracking algorithm such as recursive least-squares (RLS) or least mean squares (LMS) [34]. Due to its faster tracking capabilities, we use standard RLS in our simulations.

At a given time k, the received signal is $r_k = \underline{s}_k^T \underline{c}_k + N_k$. We assume that the transmitted symbols s_k are known at the receiver, either as training or as hard decisions supplied by the MAP detector. With the inclusion of a forgetting factor λ and initialization using an autocorrelation matrix $\delta\underline{\underline{\mathrm{I}}}$, the least-squares solution for the channel coefficients $\underline{c}_k$ is then given by:

$$\underline{c}_k = \underset{\underline{c}}{\operatorname{argmin}} \sum_{n=1}^{k} \left| r_n - \underline{s}_n^T \underline{c} \right|^2 \lambda^{k-n} + \delta\lambda^k \underline{\underline{\mathrm{I}}} \tag{2.13}$$

$$\Rightarrow \underline{c}_k = \left[\sum_{n=1}^{k} \underline{s}_n \underline{s}_n^T \lambda^{k-n} + \delta\lambda^k \underline{\underline{\mathrm{I}}} \right]^{-1} \left[\sum_{n=1}^{k} \underline{s}_n r_n \lambda^{k-n} \right] \tag{2.14}$$

The channel coefficients $\underline{c}_k$ in Equation (2.14) can be solved for recursively using the standard RLS algorithm [34]. The channel estimate obtained by

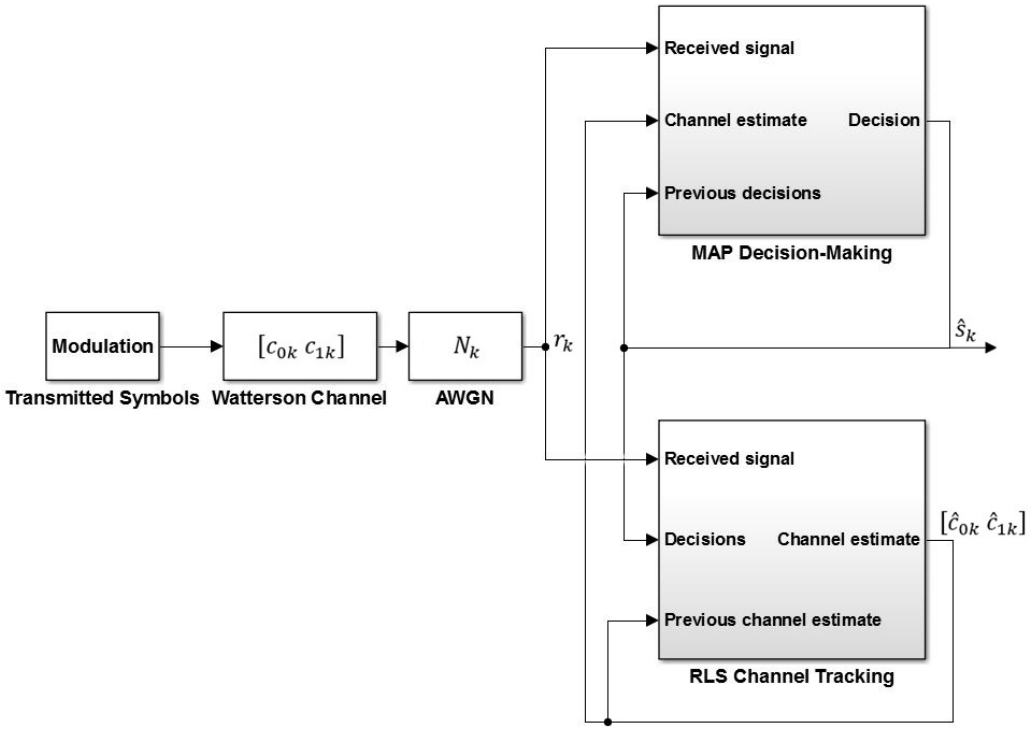

Figure 2.2: Block diagram for the coupled MAP-RLS receiver.

this algorithm is then provided to the decision-making module for future decisions.

The block diagram for the coupled MAP-RLS receiver is given in Figure 2.2. We note that this coupled receiver structure, arising naturally from the constraints of the problem, is a known receiver design philosophy, implemented in [35] and [36].

One weakness of this receiver design is the propagation of errors caused by incorrect decisions, which result in faulty channel estimates that trigger future incorrect decisions. We will observe in Chapter 4 that this receiver performs up to 5 dB worse than a MAP detector with perfect channel knowledge in our simulations. One way to improve performance is to note that an incorrect channel estimate often manifests as a dramatic change in the magnitude of the channel coefficients, indicating much more rapid variation than is plausible for the channel. This means that we can detect an incorrect decision in the past. Revising this decision and repeating the RLS update step can allow us to recover from the error event. While this correction step

can improve performance (by up to 1.5 dB in our simulations), we will obtain a greater improvement in Section 2.4 with a hybrid MLSE-MAP receiver.

We note that delaying the hard decisions made by the receiver would result in much better performance. In this case, we could evaluate potential sequences and obtain a different channel estimate for each one. If a hypothesized sequence of symbols is incorrect, then the channel estimate would diverge rapidly from the true values of the channel taps and cause further errors down the line. With a delayed decision, we could observe this divergence and choose a different sequence. The optimal receiver structure for this approach would be a form of maximum-likelihood sequence estimation (MLSE) that incorporates channel estimation. We study the design of such a receiver in the next section.

2.3 Multitrellis Adaptive Viterbi Algorithm

The use of MLSE equalizers for HF skywave communications was studied by Falconer, et al. [37], in a comparative study of MLSE performance versus that of traditional DFEs. It was observed in this paper that due to the technical constraints of the day, DFEs held significant advantages over MLSE methods for HF communications. As a result, MLSE receivers were only employed for line of sight (LOS) communications. In this section, we derive a multitrellis adaptive Viterbi algorithm (MAVA) that overcomes the obstacles to employing an MLSE for HF skywave communications. The algorithm incorporates innovations made on the traditional Viterbi algorithm in the 1990s to make it reliably adaptive, while also employing a multitrellis approach to compensate for the difficulties posed by the large tap separation that is characteristic of the HF skywave channel.

The Viterbi algorithm is a dynamic programming algorithm for calculating the smallest weight path through a trellis, which is a two-dimensional array of nodes that are connected with arcs, called branches [38]. The node in row j and column k of the trellis is denoted as node (j, k), which denotes the assertion of being in state j at time k. The incremental cost, or branch metric, of moving from state j_1 at time $k-1$ to state j_2 at time k is given by

$b_k(j_1, j_2)$. Without loss of generality, we consider the problem of finding the minimum cost path through a fully connected trellis with N states, starting from state 1 at time 1 and ending at state 1 at time M. Let the minimum cost to reach state j at time k from the known starting state be $a_k(j)$. Let the history of the best path through the trellis be stored in the array $T_k(j)$. At time k, $T_k(j)$ is the state occupied by the best path into state j at time $k-1$. The basic Viterbi algorithm can then be summarized as follows [39]:

1. Initialization: $(k = 1)$

$$a_1(j_1) = b_1(1, j_1), \tag{2.15}$$

$$T_1(j_1) = 1,$$

$$1 \leq j_1 \leq N.$$

2. Recursion: $(1 < k < M)$

$$a_k(j_1) = \min_{1 \leq j_2 \leq N} [a_{k-1}(j_2) + b_k(j_2, j_1)], \tag{2.16}$$

$$T_k(j_1) = \arg\min_{1 \leq j_2 \leq N} [a_{k-1}(j_2) + b_k(j_2, j_1)], \tag{2.17}$$

$$1 \leq j_1 \leq N.$$

3. Termination: $(k = M)$

$$a_M(1) = \min_{1 \leq j_2 \leq N} [a_{M-1}(j_2) + b_M(j_2, 1)], \tag{2.18}$$

$$T_M(1) = \arg\min_{1 \leq j_2 \leq N} [a_{M-1}(j_2) + b_M(j_2, 1)].$$

When we use the Viterbi algorithm to perform MLSE on an $L+1$-tap channel at time k, each state p at time $k-1$ corresponds to the assertion that the past L symbols were given by the sequence $\{s_p\} = \{s_p(1), \ldots s_p(L)\}$, drawn from a given constellation of Q symbols. The resulting trellis has Q^L states. The possible state transitions from state p to state q yield the sequences $\{s_{q,p}\} = \{s_q, s_p(1), \ldots s_p(L)\}$. We wish to select the transitions into each state which are optimal in an ML sense.

The inputs to the basic MLSE algorithm at time k are the received signal, r_k, and the $L+1$ channel coefficients c_i corresponding to each tap. In the absence of noise, the noise-free received signal $r_k^{NF}(q,\, p)$ corresponding to transitions from state p to state q would be given by:

$$r_k^{NF}(q,\, p) = c_0 s_q + \sum_{i=1}^{L} c_i s_p(i) \tag{2.19}$$

The branch metric $b(q,\, p)$ that corresponds to this transition is then calculated as:

$$b(q,\, p) = \left| r_k - r_k^{NF}(q,\, p) \right|^2 \tag{2.20}$$

Having defined our branch metric as in Equation (2.20), we proceed with the Viterbi algorithm to find the most likely sequence of received symbols. One way we can simplify the MLSE algorithm is to note that in practice, the most likely paths entering each state tend to converge to a single path in the past, often within several times the channel delay spread. As a result, we can truncate either our trellis or our decision-making delay to a length l in symbols that is shorter than the transmitted frame, calling this truncated length the traceback.

There are several methods to take channel variation into account for MLSE equalization. The simplest method is to obtain a channel estimate during training frames, and then consider the channel to be invariant for the decoding of data frames [40]. This approach is only feasible for the slowest-varying and most stable of channels, such as those for wired or LOS channels. The second, more advanced approach, is to use the delayed decisions of the MLSE receiver to continuously adapt to the channel. This approach is called MLSE-DD [41], and was the state-of-the-art method employed by Falconer for comparison purposes in 1985. However, the MLSE-DD strategy tracks the channel with some delay, and as a result has poor performance on rapidly time-varying channels such as the HF skywave channel. This was one of the major shortcomings of the MLSE design pointed out by Falconer. The major weakness of the standard Viterbi algorithm employed by these receivers was that only a single survivor path and metric was maintained for each state of the trellis. As a result, there was no way of quickly revising a channel estimate obtained with an incorrect decision.

The most promising strategy we currently have for MLSE equalization of time-varying channels is to update the channel estimate based on current tentative survivors for each state in the trellis, thus avoiding any delay in our channel estimate. At each step of the algorithm, we pick a number of potential paths through the trellis with the minimum accumulated metrics, designating them as survivor paths. We then perform a channel update step for each of these survivor paths, thus yielding a number of tentative channel estimates. This approach allows us to quickly fix any errors in our channel estimate. There are several related ways to implement this general idea, the most widely applicable of which are per-survivor processing [42], the generalized Viterbi algorithm (GVA) [43], [44], and list Viterbi decoding [39]. These methods mainly differ in implementation specifics, such as the retainment of a given number of survivors, or visualization and programming as a stack instead of a trellis. Seshadri's GVA approach is the one most easily and directly applicable to our problem [43].

The key weakness of MLSE receivers for HF communications is the large number of states required for the trellis, in what is known as the equalization problem for sparse ISI channels. Under disturbed conditions, the HF channel can be approximated by several taps, separated by milliseconds of delay. At high HF bandwidths on the order of 10 kHz, these taps are separated by tens of symbols, and the resulting trellis rapidly becomes computationally intractable. For example, the ITU-R midlatitude disturbed channel has a delay spread of 2 ms. At a symbol rate of 10 kbps, this is equivalent to separation by 20 taps, which translates into a trellis with 2^{20} states for BPSK symbols.

Our general solution to the MLSE HF receiver design problem is to recognize that most of the states in the trellis are independent of one another, and that we can take a much smaller trellis into account at each step. For example, in the case of the ITU-R two-tap disturbed channel at 10 kHz symbol rate and BPSK signalling, instead of a massive trellis with 2^{20} states, we can consider 20 independent trellises with two states each. At each step of the algorithm, we operate on a single trellis with its own survivor paths $\underline{\underline{T}}_k$ and associated accumulated metrics $\underline{a}_k$ using Seshadri's GVA [43]. We

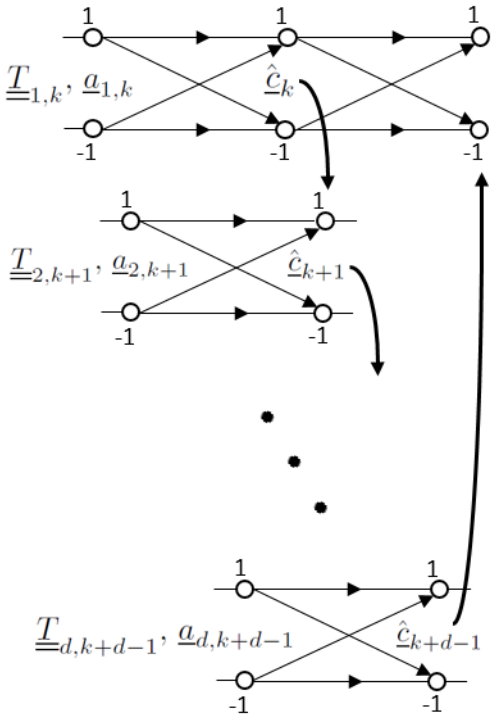

Figure 2.3: Multitrellis adaptive Viterbi algorithm operation for BPSK signaling.

update our channel estimate $\hat{\underline{c}}_k$ based on the survivor path with the lowest accumulated metric; and for the next received symbol, we use this channel estimate to calculate branch metrics for a different trellis. This process is illustrated in Figure 2.3.

In order to extend the basic Viterbi algorithm to accommodate v survivor paths for each state, we define the function $\min \mathrm{v}$, which yields the v minimum values of a discrete minimization as a sorted vector. In a multitrellis approach, the time that corresponds to the previous state will be $k-d$, not $k-1$ as with a single symbol-spaced trellis. At time $k-d$, each state j in the trellis has an associated vector of accumulated metrics $\underline{a}_{j,k-d}$ of length v, and the vector of corresponding symbols $\underline{T}_{j,k-d}$. We then modify the recursion steps in Equations (2.16) and (2.17) as follows:

$$\underline{a}_{j_1,k} = \min_{1 \le j_2 \le N,\, 1 \le u \le v} \mathrm{v} \left[\underline{a}_{j_2,k-d}(u) + b_k(j_2,\, j_1) \right], \tag{2.21}$$

$$\underline{T}_{j_1,k} = \arg\min_{1 \le j_2 \le N,\, 1 \le u \le v} \mathrm{v} \left[\underline{a}_{j_2,k-d}(u) + b_k(j_2,\, j_1) \right], \tag{2.22}$$

$$1 \leq j_1 \leq N.$$

Each surviving accumulated metric in $\underline{a}_{j_1,k}$ has an associated survivor path. Once we have obtained the vector $\underline{T}_{j_1,k}$, we use it to augment these survivor paths to obtain a matrix $\underline{\underline{T}}_{j,k}$. This matrix has dimension $(v \times l)$, where l is the traceback length.

Having determined the new survivor paths, we update the channel estimate based on the survivor path with the minimum accumulated metric, $\underline{T}_{\min,k}$. We update the channel estimate with the standard RLS algorithm, using the past channel estimate $\underline{\hat{c}}_{k-1}$ and the past sample covariance matrix $\underline{\underline{P}}_{k-1}$. Implementing an RLS function that takes the past channel estimate, sample covariance matrix, survivor path, and the received sample r_k, we can express our channel estimation step as:

$$\left[\underline{\hat{c}}_k \; \underline{\underline{P}}_k\right] = \text{RLS}\left(\underline{\hat{c}}_{k-1},\, \underline{\underline{P}}_{k-1},\, \underline{T}_{min,k},\, r_k\right) \tag{2.23}$$

There are multiple steps of sorting in the above equations, in addition to simplifications in notation due to only a single trellis being considered for updating. We provide a detailed sample-by-sample implementation of MAVA in Algorithm 1 for a two-tap ITU-R test channel, where we highlight the intermediate steps in the updating process. The extension to the case of more than two taps is straightforward, and is achieved with the use of additional states that take the right symbols into account in branch metric calculation. We note that although the traceback length and the number of survivors are variable parameters of the algorithm, for the case of the ITU-R test channels, taking a traceback of two symbols and two survivor paths per state is sufficient to achieve good performance. For a more general channel, we would expect the performance to improve with greater traceback length and more survivor paths for each state.

There are four other phases in receiver operation, which we enter as we initialize and terminate the algorithm in symbol-by-symbol processing.

1. Training: We use the standard RLS algorithm to update the channel estimate, with perfectly known training symbols.

2. Initialization: We have just left the training frame for trellis i. Before

Algorithm 1 Multitrellis adaptive Viterbi algorithm for ITU-R channels.

Parameters :

d : delay between channel taps, in symbols

l : traceback (trellis length in symbols)

$S : \{s_0\, s_1 \ldots\, s_n\}$, a symbol constellation, with $n = 2^m$, $m \in \mathbb{Z}^+$

v : number of survivor paths per state, with $v = 2^u$, $u \in \mathbb{Z}^+$.

For a two-tap channel, the number of states in a single trellis is n^2.

Given :

$\underline{\hat{c}}_{k-1} = [\hat{c}_{0,k-1}\, \hat{c}_{1,k-1}]^T$, channel estimate at time $k-1$

$\underline{\underline{P}}_{k-1}$, sample covariance matrix inverse at time $k-1$

r_k, receiver input at time k

$\underline{\underline{T}}_{i,j,k-d}$, survivor paths for trellis i, state j, and time $k-d$, $i \equiv k \bmod d$

$\underline{a}_{i,j,k-d}$, accumulated metrics for trellis i and state j at time $k-d$.

Compute :

$\underline{b}_j$, the branch metrics to be calculated for state j

$\underline{\hat{c}}_k$ and $\underline{\underline{P}}_k$, channel estimate and covariance matrix inverse at time k

$\underline{\underline{T}}_{i,j,k}$ and $\underline{a}_{i,j,k}$, survivor paths and their metrics for trellis i and state j

$\hat{s}_{k-2d}$, $2d$-delayed hard decision for a two-tap channel ($\hat{s}_{k-l}$ in general)

Computation :

1) Calculate branch metrics and accumulated metric candidates.

$\underline{\underline{T}}_{cand} = \left[\underline{\underline{T}}_{i,1,k-d} \cdots \underline{\underline{T}}_{i,n^2,k-d}\right]^T$

for $j = 1 : n^2$

for $p = 1 : n^2$

$\underline{b}_j(p) = \left|r_k - (\hat{c}_{0,k-1}s_j + \hat{c}_{1,k-1}s_p)\right|^2$ (Eq. 2.20)

$\underline{a}_{cand,i,j,k,p} = \underline{a}_{i,j,k-d} + \underline{b}_j(p) * \text{ones}(v,\, 1)$

$\underline{a}_{cand,i,j,k} = \left[\underline{a}_{cand,i,j,k,1} \cdots \underline{a}_{cand,i,j,k,n^2}\right]^T$

2) Find surviving paths; pick their metrics. Generate new survivor paths.

$\left[\underline{a}_{cand,i,j,k},\, \underline{I}_j\right] = \text{sort}\left(\underline{a}_{cand,i,j,k}\right)$; $\underline{I}_j$ a vector of sorted indices.

$\underline{a}_{i,j,k} = \underline{a}_{cand,i,j,k}(1 : v)$ (Eq. 2.21)

for $p = 1 : v$

$\underline{\underline{T}}_{temp,i,j,k}(p,\, :) = \underline{\underline{T}}_{cand}\left(\underline{I}_j(p),\, :\right)$ (Eq. 2.22)

$\underline{\underline{T}}_{i,j,k} = \left[\underline{\underline{T}}_{temp,i,j,k}(:,\, 2 : l)\ \text{ones}(v,\, 1) * s_j\right]$

3) Make decision for path with minimum metric; update channel estimate.

$\underline{m} = \left[\underline{a}_{i,1,k}(1) \ldots\, \underline{a}_{i,n^2,k}(1)\right]^T$

$[\underline{m},\, \underline{I}] = \text{sort}(\underline{m})$

switch $\underline{I}(1)$

case j

$\hat{s}_{k-2d} = \underline{\underline{T}}_{temp,i,j,k}(1,\, 1)$

$\underline{s}_{rls} = \left[s_j\ \underline{\underline{T}}_{i,j,k}(1,\, v)\right]$

$\left[\underline{\hat{c}}_k\ \underline{\underline{P}}_k\right] = \text{RLS}\left(\underline{\hat{c}}_{k-1},\, \underline{\underline{P}}_{k-1},\, \underline{s}_{rls},\, r_k\right)$ (Eq. 2.23)

carrying out the Viterbi algorithm, we initialize accumulated metrics $\underline{a}_{i,j,k-d}$ to 0. Since we have exact knowledge of past symbols, we initialize the trellis at the last known state, and only evaluate the branches coming out of this known state.

3. Survivor accumulation: We have passed the initialization step, but have not yet obtained the number of survivor paths that we want. We carry out Algorithm 1 with a few adjustments; most importantly, we do not prune the survivor paths by choosing the ones with the smallest metrics, but instead retain all new survivor paths and accumulated metrics. Once we are done with accumulating survivors, we can process the data exactly as in Algorithm 1.

4. Reentering training: While we have left the data frame and entered the training frame, we still make decisions on symbols in the data frame. During this time, we continue choosing the path with the minimum accumulated metric, while obtaining accurate channel estimates through the known training symbols.

The algorithm above makes several important assumptions. Firstly, it assumes that we have accurate knowledge of the channel profile, especially of the tap locations for the multipath arrivals, and a good initial channel estimate. These assumptions are valid if we have a long enough training frame. Furthermore, if the training frame is longer than the channel spread, we can reset the algorithm before each data frame. We can therefore start with a new set of trellises for each data frame, and thereby avoid complications due to a changing channel profile or to numerical errors resulting from running the Viterbi algorithm for too long. Secondly, the algorithm assumes that the channel consists of widely spaced taps that enable the full trellis to be decomposed into several independent trellises. If the channel consists of more than two taps with different delays between each tap, the resulting trellises will no longer be independent, and will share a number of states. While we could make the assumption of independence and proceed as above, there may be better ways of decomposing the original trellis in this case, with the dependence between different trellises taken into account. This point requires further investigation, and may lead to improved performance and a more general algorithm.

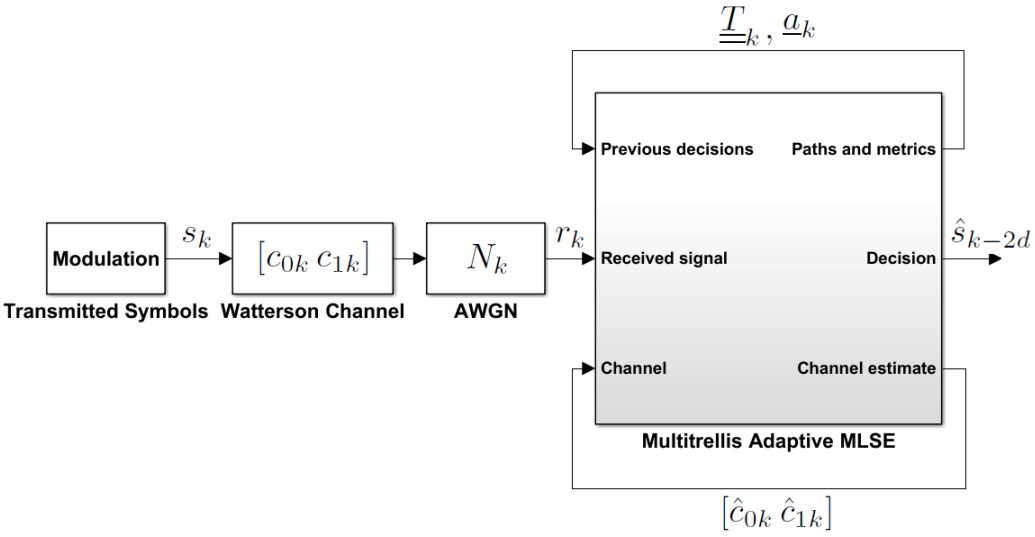

Figure 2.4: Block diagram for the MLSE receiver running the multitrellis adaptive Viterbi algorithm.

The block diagram for the MAVA receiver is given in Figure 2.4. In the next chapter, we will observe that in our simulations, the MAVA receiver leads to more than 10 dB improvement over best-case zero-delay MAP detector performance. This performance boost comes at the cost of $2d$ delay in decision-making, which is 1-4 milliseconds for the ITU-R test channels. This is a penalty that an engineer may or may not be willing to accept for a given minimum-delay communications problem. In the following section, we cascade the MAVA receiver and MAP detection to produce a receiver with zero decision-making delay that improves on the coupled MAP-RLS receiver studied previously in literature.

We note that multitrellis approaches to sparse ISI channels, and especially the HF channel, have been considered in literature. The parallel-trellis Viterbi algorithm (PTVA) [45] breaks the problem down into multiple trellises in the same fashion as our MAVA; and the multitrellis Viterbi algorithm (MVA) operates on a single trellis where only a few states are taken into consideration, based on the channel profile [46]. These methods, and the original Viterbi algorithm, are compared in [47] for complexity and performance. A multitrellis approach was independently developed for magnetic decoding applications in [48], where two parallel trellises were constructed for indepen-

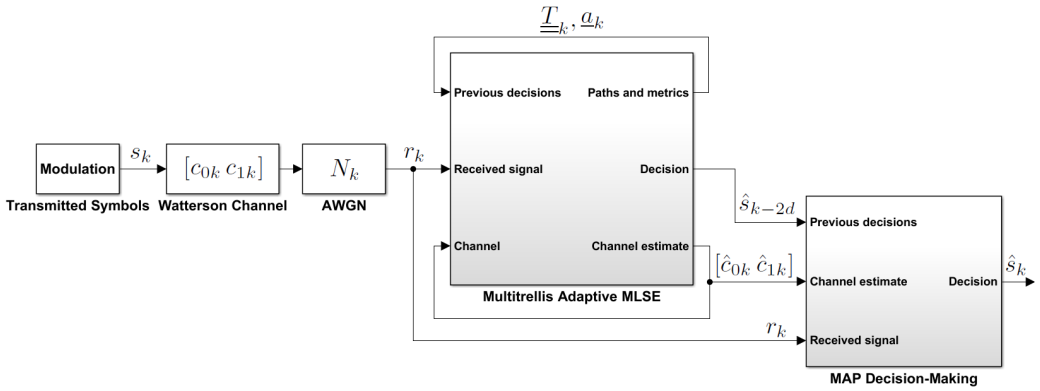

Figure 2.5: Block diagram for the hybrid MLSE-MAP receiver.

dent and interleaved dicode sequences. However, these multitrellis methods do not incorporate channel tracking into their MLSE, which has to be a key component of a practical HF receiver. To our best knowledge, the application of an adaptive multitrellis approach to HF communications is novel.

2.4 Hybrid MLSE-MAP Receiver

The first receiver we have considered in this chapter, the coupled MAP-RLS receiver, produces zero-delay decisions on the received symbols, but the coupling of decision-making and channel estimation without any delay leads to poor performance compared to MAP detection with known channel symbols. The MAVA-running MLSE receiver, on the other hand, produces more accurate decisions and a more reliable real-time channel estimate, but also enforces a delay penalty. To combine the benefits of each, we introduce a hybrid receiver where we use the accurate delayed decisions and channel estimates of MAVA as inputs to the MAP detection module. The block diagram of this receiver is given in Figure 2.5. In the next chapter, we will observe that the hybrid receiver has improved performance, and is preferable over the coupled MAP-RLS receiver.

To our best knowledge, this receiver design is novel and represents improvement over the coupled designs in literature [35] [36]. While the multitrellis adaptive MLSE block runs MAVA for our minimum-delay HF problem, this

receiver structure can be applied more generally with any MLSE implementation over a range of communication channels. It would be of interest to find applications where the coupled decision-making/channel estimation structure is used and try out the hybrid design to obtain better performance.

In this chapter, we have proposed three minimum-delay receiver designs that are directly applicable to communication over the HF skywave channel. In the following chapter, we simulate receiver performances on HF test channels and observe that all three designs are robust and exhibit good BER performance - a highly encouraging result that confirms the feasibility of minimum-delay HF communications for high-frequency trading applications.

CHAPTER 3

RESULTS

3.1 Overview

In this chapter, we present simulation results for the receiver designs discussed in the previous chapter. We evaluate receiver performance based on the BER versus SNR curves produced for the calm, moderate, and disturbed ITU-R midlatitude test channels. The simulations are performed in Simulink for 1200 seconds unless otherwise stated, at a symbol rate of 10 kbps. This corresponds to $1.2 \cdot 10^7$ symbols being transmitted. We model the channel by cascading a Rayleigh Fading block, with the appropriate parameters corresponding to the ITU-R channels, and an AWGN block that adds noise to each sample. The SNR is defined as E_s/N_0, where E_s is the signal energy in joules, referenced to 1 Ω, and N_0 is the noise power spectral density in W/Hz. Since our symbol rate is less than 12 kHz, the assumptions of the Watterson channel model are satisfied. We use a transmission frame of 120 training symbols followed by 180 data symbols. We assume that the symbols transmitted during the data frame represent the most recent data available to us, so that the training period does not impose any latency. Since we have perfect synchronization in our simulations, we do not need to employ oversampling methods, which we would use in a practical receiver.

To obtain accurate BER curves, we run 10 trials for each value of the SNR and take the average of the BER results, using different seeds each time to generate the Rayleigh fading and Gaussian noise in the channel. Due to the higher variance exhibited by the results of the ITU-R calm channel at high SNR, we run 30 trials for the SNR values that produce error rates lower than 10^{-3} for the calm channel. We use differential binary phase-shift keying (DBPSK) unless stated otherwise. Although BPSK modulation holds

an SNR advantage over DBPSK in AWGN, differential encoding is robust against phase ambiguity introduced by the channel, and would be useful in the prototyping and experimentation stages of HF modem design.

The results we present here are optimistic for the receiver designs. In real life, imperfect synchronization, simplification of the channel response, and unmodeled variations in the HF channel, such as interference bursts, would increase the BER. It is therefore advisable to deploy a more conservative and robust receiver design first, such as DBPSK with the adaptive multitrellis Viterbi algorithm, before driving up the symbol rate or using a higher order constellation.

3.2 Results for the Coupled MAP-RLS Receiver

We first simulate the performance of the coupled MAP-RLS receiver for the two-tap ITU-R test channels. We judge the performance of the receiver through comparison with MAP detection using perfectly known channel coefficients, which would yield optimal BER performance for the given receiver structure. In Figure 3.1, we observe that the MAP receiver with RLS channel tracking performs worse than MAP detection with known channel coefficients, with up to 5 dB loss on average. The BER performances of these receivers can exhibit several dB of difference between calm, moderate, and disturbed conditions for the same channel realization. On average, however, receiver performance is almost identical for different conditions, with the calm channel having the best average performance at high SNR by about 0.5 dB. The calm channel also has a higher variance than the others at high SNR, when there are fewer error events, and can be the worst channel for some trials. The reason for this phenomenon is that the rate of change of the calm channel is slow, so that within the same time interval, the receiver experiences fewer but longer lasting error events than for the moderate or disturbed channels. Therefore, a small increase or decrease in the number of such error events across different trials can have a greater proportional impact on the BER, resulting in less predictable performance.

There are several features in Figure 3.1 that are important for perfor-

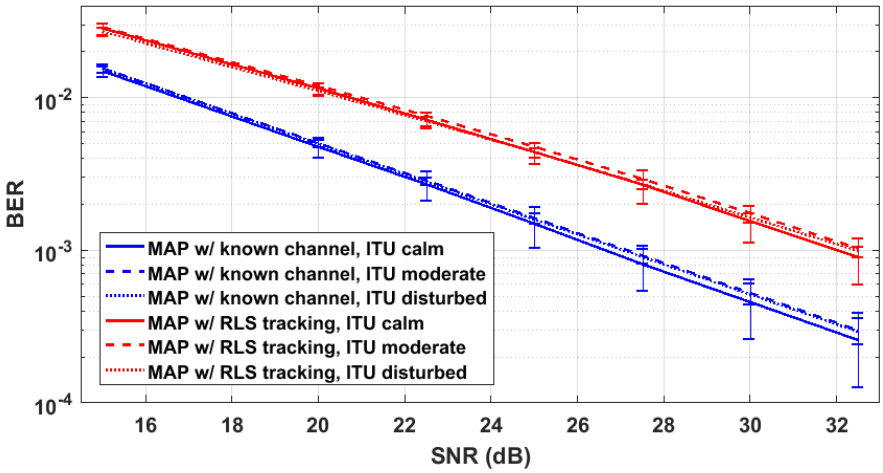

Figure 3.1: BER performance of the coupled MAP-RLS receiver, compared to MAP detection with known channel coefficients.

mance analysis. Firstly, we observe that the SNR requirement for achieving a BER of 10^{-3} - a reasonable performance goal for our target applications - is quite high. MAP detection with known channel coefficients achieves this performance at 26-28 dB, while the coupled MAP-RLS receiver requires more than 32 dB. These SNR requirements are consistent with past studies on uncoded symbol-by-symbol MAP demodulation for DBPSK signaling on Rayleigh channels [49], [50]. As a result, we can only reliably deploy a zero-delay receiver at these data rates if we have particularly good SNR.

We demonstrate the importance of keeping an up-to-date channel estimate in Figure 3.2, where we compare the BER performances of MAP detection with real-time RLS channel tracking and MAP detection with a perfectly known channel that is delayed by 50 ms. In practice, a close approximation of the past channel can be obtained using the known symbols that are relayed independently to the receiver via fiber-optic cable. We observe that the calm channel varies slowly enough that the average performance obtained using delayed coefficients is nearly identical to the case of a perfectly known channel, making it a highly serviceable channel estimation scheme under calm conditions. The performance with delayed coefficients only starts to level off after a BER lower than 10^{-3} has been achieved, resulting in a comparatively robust and serviceable receiver with minimal delay. However, the moderate and disturbed channels vary rapidly enough that use of perfect

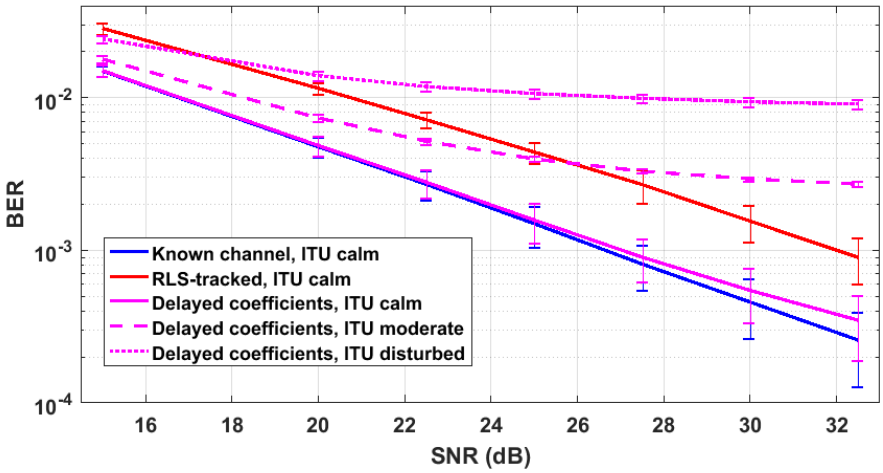

Figure 3.2: BER performance of the coupled MAP-RLS receiver, compared to MAP detection with known channel coefficients that are delayed by 50 ms.

but delayed coefficients limits BER performance to values greater than 10^{-3}. Although delayed coefficients outperform the real-time RLS-tracked estimate at low SNR, their BER curves level off before serviceable performance has been achieved. While the delayed channel estimate can be incorporated into decision-making for a calm channel, real-time tracking channel tracking is necessary to achieve robust performance across the full range of channel conditions.

In these simulations, we use an RLS forgetting factor of $\lambda = 0.925$. This relatively small λ is empirically observed to work well for the coupled MAP-RLS receiver across the range of channel conditions, because it allows the receiver to more quickly recover from faulty channel estimates caused by incorrect decision feedback. This feature of the coupled MAP-RLS receiver has been recognized in other implementations in literature [35]. In practice, the optimal forgetting factors under different conditions should be determined for the real HF channel through exhaustive experimentation.

In order to reliably achieve zero-delay decision-making for all channel conditions, we need to maintain an SNR of 30 dB or more. This can be difficult in practice, based on the state of the ionosphere at the time. As a result, we need to leverage RF engineering expertise to boost the SNR. We cover some

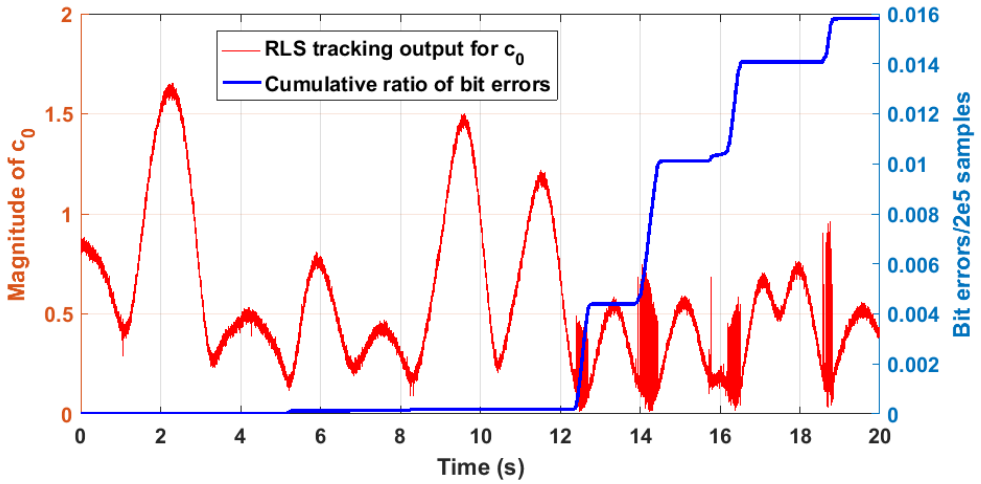

Figure 3.3: RLS-tracked channel coefficient c_0, with evolution of bit errors over time demonstrating error events during fades.

of the key design considerations for the HF link in the next chapter.

The primary reason behind the high SNR requirement for zero-delay decision-making is the inability to make tentative decisions, which could then be revised with error-correcting codes or with frame-based decoding. However, while we cannot prevent incorrect decisions from being made, we can recognize our error after the fact and revise our past decisions and channel estimates to more quickly recover from these errors. The key indicator of an incorrect decision for the coupled MAP-RLS receiver is a dramatic change in the channel estimate magnitude after the error has been made, as the estimate rapidly diverges due to incorrect decisions. This phenomenon is observed in Figure 3.3, where we plot the evolution of our channel estimate for c_0 over 20 seconds for the ITU-R moderate channel with an SNR of 20 dB. To gain insight into the factors triggering error events, we have also plotted here the cumulative number of bit errors, divided by the total number of samples over the 20-second interval. While the system recovers from incorrect estimates thanks to training, the repeated and cascading errors during the data phase are consistently the main source of bit errors.

Through our knowledge of the ionospheric channel, we observe that an overly fast rate of change in the channel estimate is an indicator of error events. Having assumed the two-tap Watterson model, we know that a sym-

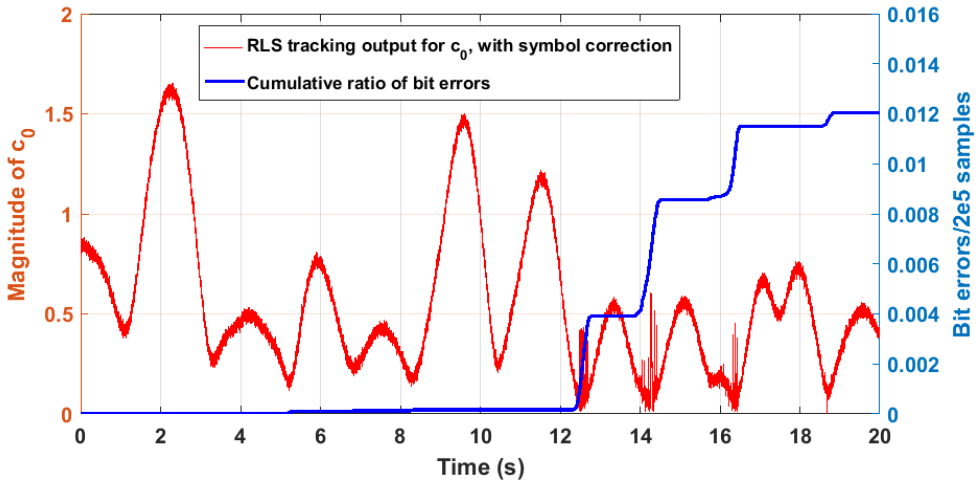

Figure 3.4: RLS-tracked channel coefficient c_0, with symbol correction used to revise our estimate.

bol error made at time k will cause an error in the channel estimate when used as an input to the RLS channel tracker. Since the same hard decision will be used at time $k + d$, we revise this decision if we observe a spike during a fade. To improve the tracking performance of our receiver, we run the RLS update with this revised decision and the current hard decision, thereby quickly recovering from the incorrect channel estimate. The potential of this approach is illustrated in Figure 3.4, where we perform the same simulation in Figure 3.3, now with this additional symbol correction step. We have empirically chosen a threshold of $\|\nabla\underline{\hat{c}}\| \geq 0.05$ for the revision step, so that we revise our previous decision and recalculate our channel estimate if we observe a faster rate of change than this. We observe that both the magnitudes and the durations of the error events are greatly reduced. In Figure 3.5, we present the BER curve obtained for RLS tracking with symbol correction on the ITU-R disturbed channel, comparing its performance to standard RLS tracking. We consistently obtain up to 1.5 dB improvement for the disturbed channel.

The key weakness of the symbol correction strategy outlined above is that it relies on the empirical selection of a threshold for the rate of change of channel coefficients. This threshold will be different across the range of channel conditions and individual realizations, and would be difficult to implement in practice due to the bursts of interference typical of HF channels. We will

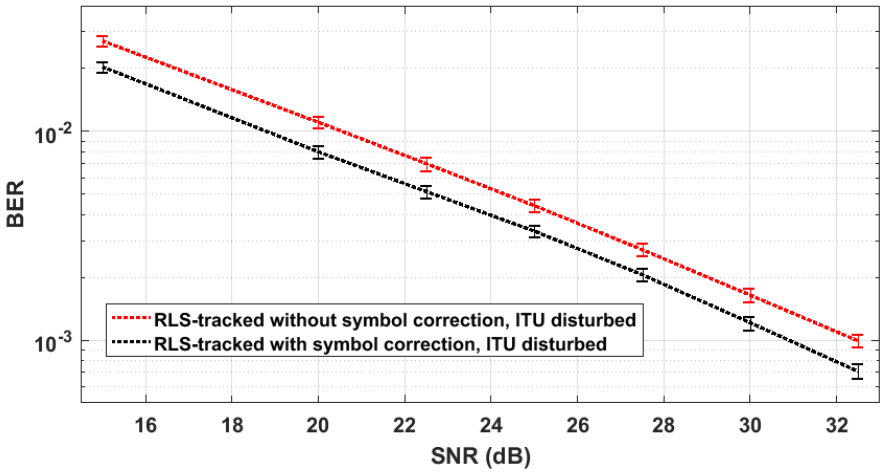

Figure 3.5: Performance of the MAP-RLS receiver, comparing RLS tracking with symbol correction to basic RLS channel tracking.

observe in Section 3.4 that the hybrid MLSE-MAP receiver will produce performance that is close to MAP detection using known channel coefficients, without resorting to optimization over specific channel conditions. It would therefore be more fruitful to focus future research on improving the hybrid receiver instead.

3.3 Results for the Multitrellis Adaptive Viterbi Algorithm

The MLSE receiver employing the multitrellis adaptive Viterbi algorithm (MAVA) allows us to delay our decisions for the data symbols, while also producing an up-to-date channel estimate. The BER curves for MAVA are given in Figure 3.6, where we compare its performance to that of MAP detection with known channel coefficients for the calm channel. We also note the respective decision-making delays of the two receivers in the figure, highlighting the tradeoff between latency and BER performance. The ability to revise our tentative decisions yields more than 10 dB improvement over zero-delay MAP detection for a calm channel. This makes MLSE a robust detection scheme that can reliably operate in much more difficult SNR regimes.

Although MAVA's current implementation yields good BER performance,

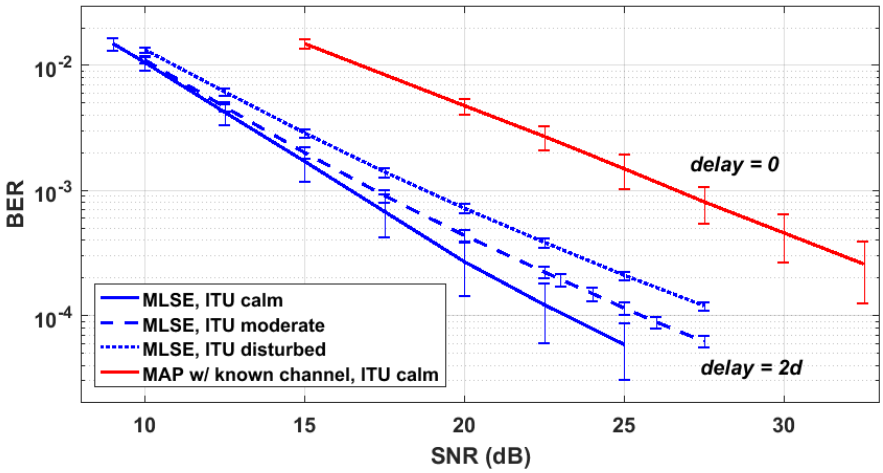

Figure 3.6: Performance of the multitrellis adaptive Viterbi algorithm, compared to MAP detection with known channel coefficients.

with decision-making delays on the order of milliseconds, it does not leverage the full potential of the generalized Viterbi algorithm [43]. Initially, we had implemented the channel estimation step exactly as in [43], by retaining and updating a channel estimate for each survivor path of each state. However, instead of dealing with a single trellis, we now have multiple parallel trellises that update the channel estimate independently of one another. As a result, we originally decided to retain the channel estimates corresponding to the survivor paths with the minimum accumulated metrics. This method yields a large number of candidate branch metrics at each trellis state that correspond to evaluations with different channel estimates. We chose the minimum branch metric for each state, and kept the corresponding channel estimates to be retained. This approach did not work in the presence of noise, where we observed that the various channel estimates were converging to different stable values instead of varying around a single correct value. We therefore chose to only retain the channel estimate corresponding to the best path at each stage, which we then pass between trellises. While the use of a single channel estimate yields sufficiently good BER performance for our application, an incorrect estimate can still cause a cascade of errors from which it is difficult to recover. We therefore need a robust method of weighting and combining the various channel estimates, so that incorrect estimates can be noticed and rectified. For the future development of MAVA, we propose to implement a variation of the soft output Viterbi algorithm (SOVA) to up-

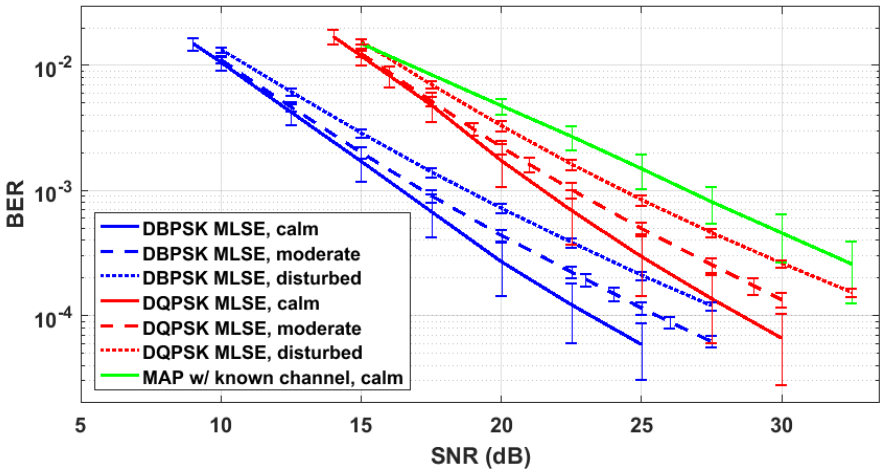

Figure 3.7: Performance for DQPSK signaling with MAVA, compared to DBPSK signaling.

date the channel estimate. We detail this possible improvement in the next chapter.

The good performance of the MAVA receiver allows us to reliably utilize a higher-order constellation over the full range of channel conditions. We provide the BER curves for DQPSK modulation using MAVA for MLSE in Figure 3.7. We observe up to 6 dB deterioration on average in performance compared to DBPSK modulation with MAVA on calm channels, but up to 7 dB improvement compared to MAP detection with known coefficients and with DBPSK modulation.

We thus encounter an interesting tradeoff between latency, bandwidth, and reliability. The use of QPSK symbols instead of BPSK causes an increase in the bit-error rate while increasing the bit rate; but additionally offers the advantage of reducing the number of symbols necessary to transmit a given message. While each symbol undergoes the same delay in detection, QPSK allows us to encode two bits into a single symbol, instead of just one bit as in BPSK. If we were simply dealing with a continuous stream of bits, higher order modulation would merely improve our data rate. However, the financial data which is of interest to the HFT user is much more likely to be available as a packet of bits, periodically updated at the data source. With a higher order constellation such as QPSK, we need half the number of symbols to

transmit a given packet of bits, and can thereby reduce the effective latency of the modem.

Suppose we have a packet of 16 bits available to transmit, and that the entire packet has to be decoded at the receiver for the information to be useful. If we employed BPSK with a symbol rate of 10 kbps for an ITU-R midlatitude moderate channel, the hybrid receiver would incur $16/10000 = 1.6$ ms of latency as it waited for the entire frame to be transmitted, but would not introduce additional decision-making delays. If we were to use MAVA with QPSK, we would introduce $2d = 2$ ms of decision-making delay, but cut the frame transmission time in half to 0.8 ms. The total latency of the MAVA receiver with QPSK is then $2 + 0.8 = 2.8$ ms; thus, we are better off using the hybrid receiver with BPSK in order to minimize latency.

Now consider the same scenario as above, but with a packet of 100 symbols instead of 16. At 10 kHz, the hybrid receiver using BPSK would have to accumulate 100 symbols over a period of 10 ms. The MLSE receiver with QPSK, on the other hand, would incur a decision-making delay of 2 ms but reduce the transmission time to 5 ms, since we now need to only transmit 50 symbols instead of 100. As a result, the latency of the MAVA-QPSK receiver becomes 7 ms, producing a latency advantage of 3 ms over the hybrid zero-delay receiver. Hence, our choice of receiver structure and modulation order is dictated by the data structures which we need to transmit for a given application. In practice, it is preferable to keep packet sizes as small as possible in order to minimize serialization delays, while encoding enough information within this small packet to successfully execute a trade.

3.4 Results for the Hybrid MLSE-MAP Receiver

By putting together MAP detection and the multitrellis MLSE receiver, we obtain performance very close to MAP detection with known channel coefficients, as observed in Figure 3.8. On average, the receiver performance is less than 0.3 dB worse than MAP detection with known coefficients for the calm and moderate channels. For the disturbed channel, the hybrid receiver exhibits up to 0.7 dB loss.

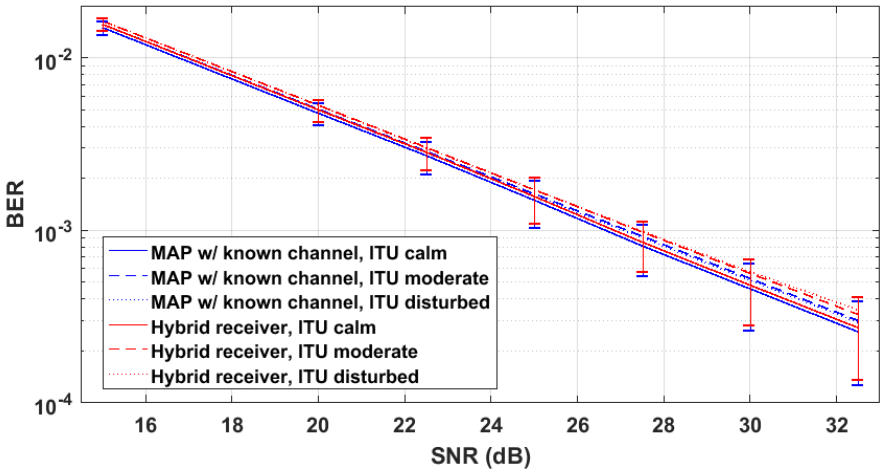

Figure 3.8: Performance of the hybrid MLSE-MAP receiver, compared to MAP detection with known channel coefficients.

We compare the performances of the hybrid receiver and the coupled MAP-RLS receiver in Figure 3.9. We observe that up to 5 dB of improvement is obtained using the hybrid receiver for calm channels, and up to 4 dB of improvement for moderate or disturbed channels. The hybrid receiver is clearly superior for zero-delay decision-making, and should be used instead of the simpler MAP-RLS receiver.

An important advantage of the hybrid design is that it incorporates both MAP detection on the received symbol, and a more accurate MLSE stage that produces a delayed but accurate decision, thus making it a flexible all-purpose receiver. Depending on channel conditions and application considerations, we can choose between the outputs of the MLSE and MAP decoders for our final decision.

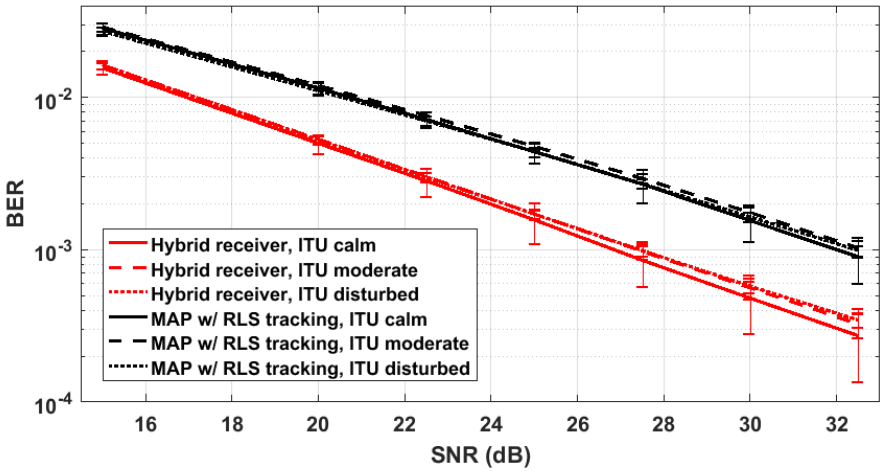

Figure 3.9: Performance of the hybrid MLSE-MAP receiver, compared to coupled MAP-RLS with symbol correction.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

In this thesis, we have studied the problem of minimum-delay HF communications. Our work was motivated by new interest in HF communications from the high-frequency trading industry, with the end goal of transatlantic transmission of financial data. Although HF communications is a well-established field in electrical engineering, the novel design constraints imposed by the trading applications led to the design of new receivers that achieved delay-optimized communication.

We started by studying the characteristics of the HF skywave channel, which arise from the physical properties of the ionosphere. Using the VOACAP and IONOLAB-RAY simulation tools, we observed that transatlantic minimum-delay HF communications were feasible for a well-designed communication link. Notably, the simulated HF link between Champaign and London had a latency of 20 ms, beating the fastest fiber connections by tens of milliseconds. Furthermore, a well-designed HF link at 10 kHz bandwidth could support more than 20 dB SNR throughout the day for 80% of the days in a month, throughout a given year. We reviewed the Watterson model for the HF skywave channel, and presented the ITU-R two-tap channel models used in our derivations and simulations. We noted that extant HF receiver designs are optimized for BER performance or bandwidth, and can add hundreds of milliseconds of latency to make the modem robust against fading and interference. Identifying the decoding and decision-making modules of the receiver as the latency bottleneck of HF modems, we focused on minimizing the delays in these stages through new receiver designs.

The first receiver that we designed coupled MAP detection with RLS channel tracking for zero decision-making delay, which we noted to be the baseline case already implemented in literature. We observed that this receiver structure suffered from error propagation that reduced performance by several dB compared to MAP detection with known channel coefficients. Next, we formulated a new multitrellis adaptive Viterbi algorithm (MAVA) that combined an adaptive Viterbi algorithm with a parallel trellis structure to solve the equalization problem for sparse ISI channels. This adaptive MLSE receiver yielded more than 10 dB improvement over MAP detection with known channel coefficients, at the cost of decision-making delays of twice the channel delay spread. Finally, we cascaded our MLSE receiver and MAP detection to obtain performance close to MAP detection with a known channel.

In our research, we have observed that even the most basic simulated HF channels are particularly difficult to work with, and force us to directly take channel structure into account in our receiver designs. To reliably communicate on these channels with minimum delay, we need to satisfy a high SNR requirement. The yearly, seasonal, and daily cycles of the ionosphere make it difficult to consistently achieve this SNR requirement. In order to continuously maintain minimum-delay transatlantic communications at a high data rate, we need to take great care in the physical construction of the HF system, using highly directive antenna arrays for both transmission and reception, supplying at least tens of kW of power for transmission, and placing the antenna arrays in areas with a low noise floor. Although the construction and maintenance of this HF communication network is a daunting challenge, it would also be an unprecedented achievement, and would be a major advance in wireless communications of great academic and commercial value.

4.2 Future Work

In our research, we have identified and circumvented the major latency bottleneck that would be encountered in HF skywave communications. However, there is a great deal of additional work that must be performed to achieve robust minimum-delay HF communications for high-frequency trading applications. Our current minimum-delay receivers are implemented in Simulink

and are designed to operate on two-tap test channels. The following steps must be taken in order to successfully establish a practical HF data link.

First, we need to simulate a practical transmitter and receiver, complete with pulse shaping, up- and down-conversion, automatic gain correction, coarse/fine frequency compensation, and phase synchronization. The modules that we have designed in this thesis would be located at the end of this transmission chain. We would use this system for offline testing of our algorithms on recordings of HF data broadcasts. These recordings would feature common issues in HF communications such as bursts of interference, colored noise, and a greater number of multipath arrivals with progressively smaller amplitudes. We could refine and expand our algorithms to correctly process these recordings.

The next step in our research would be to test our receivers in the field with HF equipment. The setup of an experimental HF data broadcasting system, especially a transatlantic one, requires a great deal of effort and know-how. Our prototype receiver would be implemented on a software-defined radio (SDR), for easy modification of our algorithms. These have a great deal of built-in functionality, such as preallocated buffering and front-end up- and down-conversion chains, which the standard user does not need to take into account. These features can be taken for granted in this prototyping stage, but should definitely be modified once a working system is demonstrated.

The final step of building the HF modem would be hardware and software optimization for minimum delay. There are performance tradeoffs involved in designing front-end filters and assigning buffer lengths: making them longer improves performance, while also increasing latency. As a result, this optimization stage should be attempted last, after a working modem link is proven to be robust across all channel conditions.

A modern transatlantic data link can be used for high-impact research in ionospheric physics and radio communications. An interesting avenue for future research is the investigation of ionospheric channel models for bandwidths of 12 to 50 kHz. The Watterson model is considered valid for bandwidths less than 12 kHz, and is thus termed a narrowband channel model.

On the other hand, wideband channel models in literature [14], [15], [16], [17] consider bandwidths of up to 1 MHz, and essentially involve a piecewise linear approximation of the channel profile across this band. Wideband channel simulators generally draw upon a large collection of soundings obtained under different channel conditions and locations to make piecewise predictions. However, this leaves an important gap in channel modeling for bandwidths of up to 50 kHz, which is too small to be considered wideband in an HF communications context. In this regime, the Watterson model may or may not apply; based on channel conditions, we may or may not have to compensate for different group delays at different parts of our bandwidth. Increasing the symbol rate up to 50 kHz reduces both the group delays of front-end filters and the total transmission time of a data frame, while driving up the SNR requirement. Furthermore, 50 kHz is a realistic upper limit on the contiguous bandwidth that can be licensed from the Federal Communications Commission (FCC) for commercial HF broadcasting, due to financial constraints and the already crowded allocation of the spectrum. Hence, accurate channel models for bandwidths between 12 and 50 kHz are of great interest. Work on these models should commence once an experimental transatlantic link has been set up and has run continuously for sufficient time to generate a large data set that accounts for most channel conditions. This channel model would be of great academic interest.

The minimum-delay receivers that we have designed are not limited to HF applications. They would be useful in any communications scenario where decision-making delays have to be minimized under difficult channel conditions. We hope to extend our research to the field of underwater acoustics, where channel conditions would be more severe, but experiments would be much easier to carry out than for HF skywave communications.

One possible method of improving our multitrellis Viterbi algorithm would be to combine it with the soft output Viterbi algorithm (SOVA) for channel estimation purposes. Currently, we only pass the channel estimate corresponding to the minimum-metric path between the parallel trellises. As a result, an incorrect decision causes error propagation as incorrect channel estimates result in a cascade of incorrect decisions. One way to mitigate this problem may be to find the best channel estimate at each state, and add

them together with exponential weighting based on the corresponding branch metrics. When a correct decision produces the right channel estimate, the contributions of the incorrect channel estimates would be negligible. If our decision is incorrect, then our channel estimate would be a weighted average of different estimates, one of which is still the right estimate. As a result, incorrect decisions would cause smaller deviations from the actual channel coefficients, and we would have a greater chance of recovering from error events.

REFERENCES

[1] R. L. Freeman, *Radio System Design for Telecommunication*. John Wiley & Sons, 2006, vol. 98.

[2] N. M. Maslin, *HF Communications: A Systems Approach*. CRC Press, 2003.

[3] K. G. Budden, *Radio Waves in the Ionosphere*. Cambridge University Press, 2009.

[4] J. M. Goodman, *Space Weather & Telecommunications*. Springer Science & Business Media, 2006, vol. 782.

[5] K. Davies, *Ionospheric Radio*. IET, 1990, no. 31.

[6] *International Audio Broadcasting for the Twenty-first Century*. National Academies, 1989.

[7] G. Lane, *Signal-to-Noise Predictions Using VOACAP - A User's Guide*. Rockwell Collins Inc, USA, 2001.

[8] E. Erdem, "Electromagnetic wave propagation model and simulation in ionosphere," Ph.D. dissertation, Hacettepe University, Ankara, Turkey, Jan. 2017.

[9] E. Erdem, F. Arikan, M. N. Deviren, and I. Cor, "A model based ray tracing algorithm for anisotropic and inhomogeneous ionosphere with GIM-TEC assimilation," in *2015 7th International Conference on Recent Advances in Space Technologies (RAST)*, Istanbul, Turkey, June 2015, pp. 477–481.

[10] E. Erdem and F. Arikan, "IONOLAB-RAY: A wave propagation algorithm for anisotropic and inhomogeneous ionosphere," accepted to be published in *Turkish Journal of Electrical Engineering and Computer Sciences*.

[11] D. Schneider, "Trading at the speed of light," *IEEE Spectrum*, vol. 48, no. 10, pp. 11–12, 2011.

[12] D. Schneider, "The microsecond market," *IEEE Spectrum*, vol. 49, no. 6, pp. 66–81, 2012.

[13] C. Watterson, J. Juroshek, and W. Bensema, "Experimental confirmation of an HF channel model," *IEEE Transactions on Communication Technology*, vol. 18, no. 6, pp. 792–803, 1970.

[14] J. A. Hoffmeyer and M. Nesenbergs, *Wideband HF Modeling and Simulation.* US Department of Commerce, National Telecommunications and Information Administration, 1987.

[15] L. E. Vogler and J. A. Hoffmeyer, *A New Approach to HF Channel Modeling and Simulation: Part II: Stochastic Model.* US Department of Commerce, National Telecommunications and Information Administration, 1990.

[16] P. A. Bello, *Wideband HF Propagation, Narrowband Interference, and Atmospheric Noise Models for Link Perfomance Evaluation.* MITRE Technical Report MT 93B0000086, July 1993.

[17] J. F. Mastrangelo et al., "A new wideband high frequency channel simulation system," *IEEE Transactions on Communications*, vol. 45, no. 1, pp. 26–34, 1997.

[18] R. Otnes, "Improved receivers for digital high frequency communications: Iterative channel estimation, equalization, and decoding (adaptive turbo equalization)," Ph.D. dissertation, Norwegian Univ. of Science and Technology, 2002.

[19] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications*, vol. 11, no. 4, pp. 360–393, 1963.

[20] J. G. Proakis, *Digital Communications*, ser. Electrical engineering series. McGraw-Hill, 2001.

[21] International Telecommunications Union, "Testing of HF modems with bandwidths of up to about 12 kHz using ionospheric channel simulators," Recommendation ITU-R F.1487, May 2000.

[22] W. N. Furman, "Robust low bit rate HF data modems," in *7th International Conference on High Frequency Radio Systems and Techniques*, Nottingham, UK, July 1997, pp. 149–153.

[23] J. W. Nieto, "Performance testing of MIL-STD-188-110B high data rate HF waveforms," in *8th International Conference on High-Frequency Radio Systems and Techniques*, Guildford, UK, July 2000, pp. 107–111.

[24] J. W. Nieto and W. N. Furman, "Improved data rate robustness of US MIL-STD-188-110C appendix D wideband HF waveforms," in *12th IET International Conference on Ionospheric Radio Systems and Techniques (IRST 2012)*, York, UK, May 2012, p. 57.

[25] S. A. Fechtel and H. Meyr, "An investigation of channel estimation and equalization techniques for moderately rapid fading HF-channels," in *IEEE International Conference on Communications, 1991. ICC'91, Conference Record*, June 1991, pp. 768–772.

[26] E. Eleftheriou and D. Falconer, "Adaptive equalization techniques for HF channels," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 238–247, 1987.

[27] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 754–767, 2002.

[28] M. Tuchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 920–952, 2011.

[29] R. Price and P. Green, "A communication technique for multipath channels," *Proceedings of the IRE*, vol. 46, no. 3, pp. 555–570, 1958.

[30] P. A. Bello, "Performance of some Rake modems over the non-disturbed wide band HF channel," in *MILCOM 88, Conference record. 21st Century Military Communications-What's Possible?*, Oct. 1988, pp. 89–95.

[31] H. V. Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. New York, NY: John Wiley & Sons, 2004, ch. 8.

[32] R. A. Iltis, "Joint estimation of PN code delay and multipath using the extended Kalman filter," *IEEE Transactions on Communications*, vol. 38, no. 10, pp. 1677–1685, 1990.

[33] K. J. Kim and R. A. Iltis, "Joint detection and channel estimation algorithms for QS-CDMA signals over time-varying channels," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 845–855, 2002.

[34] S. S. Haykin, *Adaptive Filter Theory*. Pearson Education India, 2008.

[35] L. M. Davis et al., "Extended least squares identification of doubly spread mobile communication channels," in *Proc. Int. Conf. Telecommunications*, Melbourne, Australia, Apr. 1997, pp. 423–428.

[36] K. Fukawa and H. Suzuki, "Adaptive equalization with RLS-MLSE for frequency-selective fast fading mobile radio channels," in *Global Telecommunications Conference, GLOBECOM '91*, Phoenix, AZ, USA, Dec. 1991, pp. 548–552.

[37] D. Falconer et al., "Comparison of DFE and MLSE receiver performance on HF channels," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 484–486, 1985.

[38] R. Cypher and C. B. Shung, "Generalized trace back techniques for survivor memory management in the Viterbi algorithm," in *Global Telecommunications Conference, 1990, and Exhibition. 'Communications: Connecting the Future', GLOBECOM'90, IEEE*, Dec. 1990, pp. 1318–1322.

[39] N. Seshadri and C. E. W. Sundberg, "List Viterbi decoding algorithms with applications," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 313–323, 1994.

[40] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, 1972.

[41] J. Lin, F. Ling, and J. G. Proakis, "Joint data and channel estimation for TDMA mobile channels," *International Journal of Wireless Information Networks*, vol. 1, no. 4, pp. 229–238, 1994.

[42] R. Riccardo, A. Polydoros, and C. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 354–364, 1995.

[43] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 1000–1011, 1994.

[44] J. G. Proakis, *Adaptive algorithms for blind channel equalization.* New York, NY: Springer, 1995, pp. 139–151.

[45] N. C. McGinty, R. A. Kennedy, and P. Hocher, "Parallel trellis Viterbi algorithm for sparse channels," *IEEE Communication Letters*, vol. 2, no. 5, pp. 143–145, 1998.

[46] N. Benvenuto and R. Marchesani, "The Viterbi algorithm for sparse channels," *IEEE Transactions on Communications*, vol. 44, no. 3, pp. 287–289, 1996.

[47] J. Mietzner et al., "Trellis-based equalization for sparse ISI channels revisited," in *Proc. International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005, pp. 229–233.

[48] R. D. Cideciyan et al., "A PRML system for digital magnetic recording," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 1, pp. 38–56, 1992.

[49] M. Gertsman and J. H. Lodge, "Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels," *IEEE Transactions on Communications*, vol. 45, no. 7, pp. 788–799, 1997.

[50] C. D'Amours, M. Moher, and A. Yongacoglu, "Comparison of pilot symbol-assisted and differentially detected BPSK for DS-CDMA systems employing RAKE receivers in Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, pp. 1258–1267, 1998.