Robust Ultra High Frequency (UHF) Satellite Communications Protocol for UUVs

DoD SBIR Number 02.1, Topic N02-019
Contract Number N66604-02-C-4577

Final Technical Report

Contract Line Item Number 0001AE
Wavix Shipment Number WAV005

Prepared by:

Dr. Jeffrey N. Shaumeyer
Principal Investigator
John Borden, President
Wavix, Incorporated
6435 Virginia Manor Road
Beltsville, MD 20705
Phone: (240) 264-8260
Fax: (240) 264-8263
E-mail: jborden@wavix.com

Issuing Government Activity:

Naval Undersea Warfare Center Division, Newport
Code 5911, Simonpietri Drive
Newport, RI 02841-1708

Security Classification

UNCLASSIFIED

Table of Contents

List of Figures

A. Executive Summary

Wavix has developed a recommendation for an optimized-protocol solution to the problem of RF satellite communications in disadvantaged marine environments. Our solution invokes a diverse but coordinated array of noise-mitigation techniques that function at the lowest levels of the 7-layer OSI Model, namely, the Physical and Link Layers. Our requirements analysis has gone beyond just noise characteristics. We have also considered constraints imposed by the physical limitations of the application, e.g., small UUVs or profiling buoys of limited power, weight, and size, as well as compatibility with existing UHF SATCOM systems.

We have looked at several effects of water and seastate on RF propagation and their likely impact in the maritime environment. The two primary concerns are signal attenuation due to waves that obscure the line of sight, and channel fading due to multipath interference caused by reflections off the rough sea surface. The mechanisms are different, but their effects on the signal channel are similar, predominantly at lower-elevation look angles. At UHF frequencies, rain and atmospheric-moisture attenuation is a very weak effect and can be neglected in comparison to other effects.

In considering mitigation strategies, we have been looking primarily at characteristic disturbance or fade times (i.e., wave periods) and their relation to wave height. The characteristic times associated with significant wave heights of 1 to 3 meters (i.e., sea-state 4) is on the order of 5 to 10 seconds. A fading channel with a characteristic time near 10 seconds poses several challenges. Given that we are considering a data link, rather than a real-time voice link, there are coding techniques that offer decided advantages. Long time-scale interleaving, when used with forward-error correction techniques, can provide a level of immunity, but at a cost of increasing latency in the signal transmission. We have also considered compatible error-reduction strategies and approaches to embedding higher-level protocols that can further increase data-transmission efficiency. We have made a thorough survey of existing channel-coding and forward-error correction techniques and their applicability.

Working from system requirements, we have performed a broad-ranging survey of existing and emerging network systems, as we seek to identify state-of-the-art concepts and system constraints that would contribute to our protocol-design effort.

B. Summary of the Effort

1.) Technical Objective #1: Characterize the Marine Environment

a.) Original Description of Objective

We want to develop a thorough understanding of the marine-noise environment. We will survey relevant research literature to establish spectral characteristics of ocean waves and understand their effects on RF signal propagation. We will also investigate multipath interference effects caused by the varying ocean surface for its impact on signal discrimination. This survey will es-

tablish characteristic wave times and their relationship to time to recapture the bitstream after fading or block error events and establish the statistical noise profile.

b.) Results

We have located a fair amount of useful real-time and historic (i.e., long-time averaged) data from several sources: the National Data Buoy Center, the (DoD/Navy) Fleet Numerical Meteorology and Oceanography Center, and the (Canadian) Marine Environmental Data Service.

Our analysis of some of these data points out that the characteristic times associated with significant wave heights of 1 to 3 meters (i.e., sea-state 4) is on the order of 10 seconds. The following are some characteristic length scales in this environment:

- Wave Heights: 2 m
- RF Wavelength: 1 m (@ 300 MHz)
- Surface Roughness: 0.1 m

When the length scales of two possibly interacting phenomena are widely dissimilar, the phenomena are unlikely to interact; the possibility for strong interaction increases when the length scales are similar. We see from this that, while the ocean waves are close to the RF wave lengths, we can expect very little interference resulting from surface roughness.

As with length scales, when time scales differ widely, interaction tends to be minimized. When time scales are similar, and there is a method of interaction, interaction tends to be maximized. The following are some characteristic time scales in this environment:

- Wave times: ~5-10 s
- SATCOM frames:
    - 8.96 s, made of 1024 blocks [5 kHz waveform]
    - 1.3866 s, made of 0.052 ms time chips [25 kHz waveform]
- Geo propagation time: 500 ms
- SATCOM blocks:
    - 8.75 ms blocks in 5 kHz waveform; variable number/channel
    - 0.052 ms “time chips” in 25 kHz waveform; channel format varies but order of 100 might be typical
- ATM packet size: ~2.5 ms
    - IP packets are 4 to 10 times larger, but variable
- Symbol size: ~0.05 ms (@ 19.6 bits/second)

The most significant information provided here is the wave period. This gives us some useful assumptions about the nature of signal fading from wave shadowing. If the wave period is 5 to 10 seconds, then we can infer that the UUV will be in the trough of the waves for 1/3 of that time or 1 to 3 seconds. The duration of any resulting loss of signal will depend on the elevation of the satellite and the height of the UUV antenna off the water surface. Therefore, we must be able to recover from a periodic loss of signal with a duration of 1 to 3 seconds and with a period of 5 to 10 seconds. This information will be key to choosing the optimal size for the bit interleaver.

2.) Technical Objective #2: Evaluate Existing UHF SATCOM Compatibility

a.) Original Description of Objective

We will survey the implementation of existing UHF SATCOM systems with several goals in mind: 1)Identify additional system requirements; 2) Evaluate potential compatibility between new protocols and existing systems; and 3) Evaluate the effect on implementation suggestions.

b.) Results

This task blossomed into a broad-ranging survey of existing and emerging network systems, as we sought to identify state-of-the-art concepts and system constraints that would contribute to our protocol-design effort. Given the strong mandate for interoperability, these technologies all impinge in one way or another.

- Cellular Telephone & PCS: a possible UUV/Satellite network shares many architectural similarities with Cellular networks (think of the satellites as mobile cell towers), but with important differences due largely to the greatly increased distance across the air-link (power and latency are primary), and to the nature of the traffic carried by the network (Cellular is a circuit-switched network, but the UUV/Satellite system would be packet-switched or message-switched). The three air-link techniques used in PCS, TDMA, GMS, and CDMA, are all potentially applicable to our protocol consideration. Although TDMA is considered old fashioned technology in the Cellular world, our finding is that TDMA is likely still the most suitable to the disadvantaged marine RF environment.
- Internet & ATM Networking: Interoperability with the Internet is strongly desirable. Although the TCP/IP suite of protocols is not directly suitable for satellite networking, the Internet is a vast networking experiment backed by a storehouse of knowledge. ATM is important since it contributes state-of-the-art concepts about packet-management in a packet-switched network. Some of the concepts in our proposal are drawn from an ATM networking environment.
- The Interplanetary Internet: NASA is in the early concept-development stages that will lead to what it calls an "Interplanetary Internet", that could build a network from a heterogeneous array of manned and unmanned orbital and interplanetary devices. The Interplanetary Internet seems to be the only research effort that is addressing two problems relevant to our own efforts: dealing with high latency (sometimes exceedingly high latency) in datagram propagation, and dealing with non-persistent network connections and routing. It is in this context that considering a message-switched approach to network operation gains value.
- Wireless Networking: emerging standards (IEEE 802.11, Bluetooth, some aspects of WAP) are generally concerned with short-haul interconnection between wireless devices, with an emphasis on accommodating Internet connectivity. The protocol details are not specifically applicable to our work, but these protocols encapsulate state-of-the-art under-

standing about user authentication and network security. In addition, they provide useful insight into protocol mechanisms for negotiating virtual connections between devices.

i.) UHF SATCOM DAMA Compatibility

We reviewed MIL-STD-188-181 (single-access UHF satcom), -182 (5-kHz UHF/DAMA waveform) and -183 (25-kHz UHF/DAMA waveform). Since our initial concept for improved communications protocols suggested a Time-Division, Multiple-Access approach, and since UHF/Satcom is a TDMA system, there is top-level compatibility in our approaches. We have found that both DAMA waveforms offer a usable system in which we can implement some additional noise-mitigation techniques that could offer significant coding gains. Below, we outline a trade study in which we show which techniques could be usefully implemented within existing UHF/DAMA specifications, which would retain system compatibility on a single-use channel, and what might be possible if the communications channel went beyond strict 188-181/182/183 compliance. A significant issue in the implementation would be maintaining interoperability while increasing the RF channel quality without undue radio complexity.

ii.) Existing UHF SATCOM Radio Hardware

We have also carried out an informal survey of existing UHF SATCOM radio hardware. We found that commercially available SATCOM radios have tended towards the more complex and more expensive for the past few years. The large military programs have demanded multi-functional radios that can utilize all the capabilities of the UHF SATCOM system and even be able to operate on other military satellite systems. This has resulted in the end of production for simple, inexpensive radios of the type needed for the UUV program. The added capability of these radios results in hardware that is much larger and heavier than necessary and uses more power than needed. In addition, we have determined that some of our protocol recommendations are not likely to be supported in these off-the-shelf radios. It is our recommendation that a new radio be designed with the single function of data transfer in the marine environment. This would result in hardware that is smaller and lighter, minimizes power consumption, and is optimized for automated operation in the disadvantaged marine environment. It would also be much less expensive, perhaps by as much as an order of magnitude.

3.) Technical Objective #3: Evaluate System-Use Constraints

a.) Original Description of Objective

This task looks at system constraints from the user viewpoint. Evaluations include: 1) Likely user density; 2) Desirable and practical data rates; and 3) Most likely embedded protocols used.

b.) Results

We used existing UHF/SATCOM requirements as a conceptual baseline against which to compare our preliminary ideas and as an aid in making our recommendations.

i.) Typical User Distribution

The issue behind this question has to do with geographical user density: how many users will simultaneously desire to communicate with the satellite at any given moment. Any satellite communication system has limits to the number of simultaneous users that can be accommodated. This is handled by some combination of concurrent communication channels and service queuing. There are many different ways to deal with this issue including simple queuing based on first-come-first-served, prioritized access, time-division multiple access (TDMA), code-division multiple access (CDMA), multiple transponders and frequencies on each satellite Frequency-Division, Multiple Access, or FDMA), additional satellites, and establishing artificial limits for each user.

In oceanography, users are widely disbursed across the world so the average density is very low. The Argo Program is useful as an example. A typical low-Earth orbiting satellite has a footprint of about 18.7 x 10^6 km^2. The Argo Program has a goal of placing drifting buoys on a 3º-square grid, giving an average area of 0.11 x 10^6 km^2 occupied by each float. So, there would be about 170 floats/footprint, a not impractical number to service.

On the other hand, many oceanographic and meteorological programs do not disburse their platforms widely, rather they tend to be stationed along coastlines or other areas of particular interest. A typical low earth orbiting satellite footprint, as noted above, easily encompasses the entire east coast or west coast of the US and Canada, or all of Australia and New Zealand, etc. Along the east coast of the US and Canada, there are as many as 1,000 reporting platforms of various types that are operated by the National Data Buoy Center and/or the Meteorological Service of Canada. If we include other types of users such as the Voluntary Observing Ship (VOS) program, the National Water Level Observation Network (NWLON), the Automated Surface Observing Systems (ASOS), and the various drifting buoy programs, the number of potential simultaneous users within a satellite footprint quickly balloons to many thousands.

ii.) Desirable and practical data rates

The data rate on the links to the UUVs is a trade-off between power requirement and time to upload data. Higher data rates require more power, but less transmission time. It is also the case that the higher the data rate, the more frequency bandwidth per channel is required. The frequency bandwidth allocation should be sufficient to allow for as many channels as possible to accommodate the number of potential simultaneous users discussed above. The necessary bandwidth for a single channel is a function of the modulation scheme, data rate, deviation, and Doppler correction. Using a combination of TDMA channel sharing and throughput limits would easily accommodate a very high number of simultaneous users while allowing throughput limits well beyond current industry expectations.

iii.) Most likely embedded protocols

There are a number of potential modulation schemes, some of which use less bandwidth or less power for equivalent data rates. Other modulation schemes are better at rejecting interference but use more bandwidth. Below, we discuss in detail the most efficient modulation techniques to minimize power use and maximize data rate.

4.) Technical Objective #4: Identify System Requirements

a.) Original Description of Objective

This task draws together the inputs from all the previous tasks into a statement of relevant system requirements as they affect the design of the protocol stack. In addition, other system requirements will be identified. Additional considerations will include: 1) Power Requirements; 2) Typical Link Budgets; 3) Antenna Design; and 4) Near-Water Operation

b.) Results

i.) Maritime RF Noise Sources

We looked at several effects of water on RF propagation and their likely impact in the maritime environment. In our proposal we identified six likely maritime noise sources:

- Signal fading from wave motions rocking UUV
- High seas coating antenna or changing local RF propagation characteristics
- Shadowing by waves
- Surface wave reflections causing multi-path interference
- Changing atmospheric conditions in LOS
- Nature of channel noise: high BER?

Of these issues, we have determined that certain of them are non-obstacles. Below we identify the non-obstacles and outline the reasons we have discounted them; these are discussed further in a later section.

- Antenna wash-over
  - Choose appropriate non-conductive housing
  - Water has small attenuation at UHF
- Ground-plane issues
  - Appropriate choice of antenna:
    - Quadrifilar Helix has small sensitivity to ground plane
    - Helical antenna is very sensitive to ground plane
- Water aerosols
  - At UHF frequencies, rain and atmospheric-moisture attenuation is a very weak effect and can be neglected in comparison to other effects
  - (Atmospheric water effects become serious above ~2-4 GHz)
- UUV rocking
  - Motions are too small to disrupt signals
  - Motions are too slow for Doppler effects

Two significant wave- and water-related issues remain. One is signal attenuation through waves that obscure the line of sight between the UUV and the satellite; the other is channel fading due to multipath interference caused by reflections off the rough sea surface. These are the significant contributors to signal degradation in the maritime environment. The mechanisms are different, but their effects on the signal channel are similar, predominantly at lower-elevation look angles. In considering mitigation strategies, we have been looking primarily at characteristic disturbance or fade times (i.e., wave periods) and their relation to wave height.

ii.) Satellite Elevation vs. Latitude

- Waves Obscuring LOS
- Fading by attenuation at small look angles; some multipath at large look angles
- Operations in northern latitudes: near-horizon viewing of SATCOM satellites
- Operations in sea-state four: wave heights of 1.25 to 2.5 meters
- Depth of fades: ~7 dB @ 1 GHz
- Wave period: 5 to 10 seconds
- Time scale of fades: ~1-3 seconds

iii.) Ocean Surface Roughness

- Fading effects from multi-path interference
- Length scales of ~0.1 meter mean less interaction with UHF signals
- Most serious at very small look angles

iv.) Nature of the RF Channel

- The RF channel is principally a fading channel (Rayleigh channel, or channel with memory), and not a noisy channel (AWGN channel, or memoryless channel)
- BER has limited utility
- Memory channels are much more difficult to simulate (Markov chains)
- Much less research exists for fading channels

A fading channel with a characteristic time near 10 seconds poses several challenges. However, given that we are considering a data link, rather than a real-time voice link, there are coding techniques that offer some hope. Long time-scale interleaving, when used with forward-error correction techniques, can provide a level of immunity, but at a cost of increasing latency in the signal transmission. We'd like to refine our understanding of the ocean-surface model to find the right balance in the trade off.

v.) Multipath Interference Effects

We have identified in the literature a model [discussed in *Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems*, by Julius Goldhirsh and Wolfhard J. Vogel, chapter 9; reference 5] for maritime multipath interference effects at L-band frequencies. ITU recommendations adopted the model in a frequency range of 0.8-8 GHz ["Propagation Data Required for the Design of Earth-Space Maritime Mobile Telecommunication Systems", Recommendation ITU-R P.680-3; reference 12]. We would like to extend those results to our UHF frequencies of interest, but that would represent a significant research effort. Nevertheless, we believe that we will be able to use current model results to provide some reasonable bounds on the magnitude of multipath-loss effects.

vi.) Antenna Design

With some antenna designs, ground-plane effects are important for operation and there is some concern about maritime operation where the ocean surface, which would function as the ground plane, is subject to disruption through swell and sea. We will recommend that these concerns are best solved by using an antenna design like a quadrifilar helix whose gain pattern does not rely on a ground plane.

5.) Technical Objective #5: Develop the Detailed Concept for the Protocols

a.) Original Description of Objective

This task proceeds from the system requirements resulting from the previous task. We will put together a complete and detailed concept of the low-level protocol stack.

b.) Results

A high-level outline for the protocols is adequately provided by the existing UHF/SATCOM 5-khz and 25-khz waveforms. With that starting point, there is a range of options for error mitigation that can be characterized by:

- Coding gain (i.e., effective increase in signal/noise ratio)
- Implementation benefits and penalties
- Compatibility & interoperability with existing UHF/Satcom systems

We have made a thorough survey of existing channel-coding and forward-error correction techniques and are determining their applicability according to the criteria above. The degree of interoperability with existing systems is the primary constraint on which techniques may be used and at what level in the protocol hierarchy. For the final report we are preparing a trade-off list that can be used as the basis for determining a final approach.

6.) Technical Objective #6: Suggest Implementation Approaches

a.) Original Description of Objective

The results of the previous task will be summarized by a presentation of suggested implementation concept designs, with an evaluation of associated hardware and software requirements.

b.) Results

Reported below.

C. Research Results

1.) Statement of the Problem

The problem is to be able to establish reliable, real-time communication, via UHF SATCOM satellites, between an unmanned underwater vehicle (UUV) and its controlling station. The UUV may be in any maritime location worldwide, experiencing up to sea state 4.

In our proposal, we listed the following as possible impediments to reliable communication:

- Signal fading from wave motions rocking the UUV;
- High seas coating the antenna or changing the local RF propagation characteristics;
- Shadowing by waves causing fading in the UHF signal;
- Surface wave reflections causing multi-path interference in the UHF signal;
- Changing atmospheric conditions in line-of-sight between the UUV and the satellite; and,
- Nature of channel noise in the maritime environment possibly leading to unusually high bit-error rates.

During our studies we refined the list into the following eight topics:

- Antenna Wash-Over
- Variable Ground-Plane
- Water Aerosols
- Rocking motions of the UUV
- Small look angle to satellite
- Wave obscuring
- Ocean-surface roughness
- Nature of the RF channel

The first four topics we believe are minor concerns that are easily dealt with; they are discussed briefly in the next sections. The remaining four are more serious concerns at the heart of our research, and will be addressed in greater detail.

2.) Minor Concerns

a.) Antenna Wash-Over

With the radio antenna necessarily near the water's surface, it will frequently be washed over and coated by seawater. Certainly waves passing over the antenna are likely to cause signal fading, but the concern is that the radio link might be disrupted by attenuation from a thin layer of seawater coating the antenna, or by water grounding the antenna.

The antenna should be adequately protected from electrical grounding if it is encased in an appropriate, non-conductive housing that is transparent to UHF wavelengths. This housing may then become coated with a layer of seawater, but seawater attenuation at UHF frequencies is small enough (< 3 dB/m) that it should not be a serious problem.

b.) Variable Ground Plane

With many possible designs of UHF antennas for UUVs, the surface of the water could be used as a ground plane. The ground plane alters the antenna-gain pattern so that its gain is preferentially above the horizon. The concern is that variations in the surface of the sea near the communicating UUV could cause variations in the antenna gain and disrupt UHF communications.

This could indeed be a problem, although it is worth noting that the antenna is sensitive to the ground plane only within about one wavelength. Some antenna designs, e.g. a helical antenna or turnstile antenna, depend on the ground plane to create a suitable gain pattern. However, a quadrifilar helix antenna is largely insensitive to the ground plane, and would be suitable choice for use on the UUV.

- ☑ Use a quadrifilar helix antenna, which is relatively insensitive to the ground plane, to avoid concerns about a variable sea surface.

c.) Water Aerosols

The immediate maritime environment of the UUV, particularly in high seas, could contain a high density of water aerosols in the form of spray, mists, and the like. The concern is that these water aerosols might significantly attenuate the UHF signal and disrupt UHF communications.

At UHF frequencies, attenuation by water is very small (< 3 dB/m), and water aerosols, even relatively dense mists, will have a negligible effect. The same is true for atmospheric moisture (fog, clouds, rain, etc.), whose effects are not significant for frequencies below several gigahertz.

d.) Rocking Motions of the UUV

Wave motions in the UUV location will cause the UUV to undulate and rock; the concern is that these rocking motions might cause fading of the UHF signal and disrupt communications.

We believe that there will be negligible effect on signal quality due to wave-induced motions of the UUV. The wavelength of the UHF signal is roughly one meter, and one "bit" of digital signal (at, say, 10,000 bits/s) is transmitted in 100 μs. From these length- and time-scale considerations, we see that rocking motions of the UUV are both too small and too slow to change significantly during the transmission of a signal bit, thus unable to disrupt signal reception significantly. Put another way, the possible motions are too slow to cause any noticeable Doppler shifts.

3.) Major Concerns

In the following sections we discuss the main impediments to robust maritime communications, and the strategies that will mitigate those impediments. The major remaining concerns are:

- Small look angle to satellite
- Wave obscuring
- Ocean-surface roughness
- Nature of the RF channel

The first is a serious issue that is not easily addressed by protocol considerations; we discuss it in the next section. The remaining concerns are the ones that implicitly determine our approach to selecting protocols for UUV communications; their nature and implications are explored in the remainder of this report.

a.) Satellite Elevation

A clear, unobstructed line of sight between a UUV and its server satellite is necessary for reliable communications. All existing SATCOM satellites (and leased services) are geostationary satellites, and therefore have orbital latitudes of 0°. Considering the desire for worldwide availability of communications to UUVs, this poses a potentially serious geometrical problem that cannot be entirely overcome with improved communication protocols.

Maintaining a reliable link with a communications satellite is difficult when the elevation of the satellite is low (where low means within 15° to 20° of the horizon), whether from increased atmospheric effects (due to greatly increased path-length through the atmosphere at low elevations), enhanced multi-path effects, more likely obstructions to the line of sight, or (for the case of low-earth orbiting satellites) greatly increased range to the satellite. For a UUV in sea-state-4 conditions, waves can be up to 2.5 m in height with a wavelength of approximately 150 m (a period of roughly 10 s); these waves will obscure everything between the horizon and 15° elevation between 5% and 10% of the time, and regions below 40° between 1% and 2% of the time.

The elevation, E, of a geostationary satellite is given by the following equation:

$$E = \arctan\left[\frac{\cos(G)\cos(L) - 0.1512}{\sqrt{1-\cos^2(G)\cos^2(L)}}\right] ,$$

where L is the latitude of the UUV, and G is the difference between the UUV longitude and the satellite longitude. Figure 1 plots E for several values of G.

Even for a client in the same longitude as the satellite, elevation falls below 40° above latitudes of 50°, roughly the latitude in the northern hemisphere of the US-Canadian border, or the English Channel. This may not be a problem for many users, but it is an impediment to global readiness.

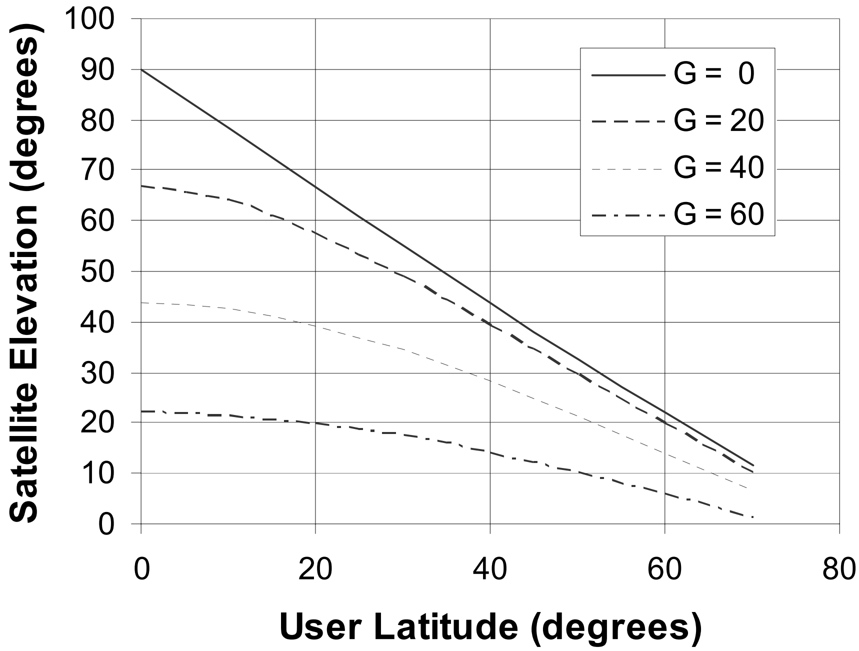


Figure 1: Satellite Elevation vs. Client Latitude

When we discuss strategies for mitigating the problem of wave obscuring, they will also address the problem of low satellite elevations, but protocol considerations will have a limited utility, in effect increasing somewhat the latitude at which geostationary SATCOM satellites can be used, but there will still be significant high-latitudes areas that will be excluded from reliable communications. The combination of waves large enough to obscure the line-of-sight to the satellite, and high latitudes that place a geostationary satellite low on the horizon is probably the most difficult challenge to reliable, global SATCOM use by UUVs.

The only satisfactory solution for high-latitudes is to use a constellation of non-geostationary satellite (i.e., those with low-earth orbits, medium-earth orbits, or highly eccentric orbits). We note in passing that using a non-stationary satellite system would be easier to implement for UUV users, since the UUV will necessarily use an omni-directional antenna, whereas most current SATCOM clients use directional antennas unsuitable for use with a non-stationary satellite network.

b.) The Effects of Waves

The effects of waves on UHF communication is two-fold: from obscuring of the line of sight to the satellite, and from reflections and scattering that might cause multi-path interference in the signal.

In sea state 4, significant wave heights are 15. to 2.5 meters with periods of about 8 s. The approximate dispersion relation for deep-sea gravity waves is

$$\omega^2 = gk \quad ,$$

where $k = 2\pi/L$ is the wavenumber (L is the wavelength), and $\omega = 2\pi/T$ is the angular frequency (T is the period). Thus, waves with 8-second periods have a wavelength near 100 m.

Wave heights of 2.5 m will obscure a satellite at 15° elevation for blocks of time that are about 9% of the wave period (wave-height/[tan(15°) × wavelength] = 0.09), which is a block of time some 0.75 s long; the block of time is about half that for an elevation of 30°. Thus, the characteristic time for block noise in the communications channel to the order of 1 s.

The depth of the signal fading will depend on the wave height (hence the transmission path length through the wave) and is difficult to estimate. Results are reported by reference [5], but are for radio frequencies only above 1 GHz. Extrapolating those results to UHF and using the result as an upper limit suggests that we should expect fading depths of as much as 7 dB.

Multi-path effects from reflections and scattering from ocean-surface roughness will also be present, and will also cause fading in the signal. Surface roughness is generally on a shorter length scale of 0.1 m or less, and is much less a problem for UHF signals than at higher frequencies; what multipath interference there is will be most troublesome at low angles of incidence, i.e., when the satellite is at low elevation. The same is true for multipath interference that result from reflections from larger waves.

We conclude that all of these obstacles can be addressed by looking for protocol strategies that mitigate the problem of signal fades to depths of 7 dB that last on the order of 1 second, with a period of roughly 8 seconds.

c.) The Nature of the Communication Channel

Noisy RF channels are classified in two ways: as Additive White Gaussian Noise (AWGN) channels, if the noise on the channel is randomly distributed, or as Rayleigh Channels if the noise arrives in blocks.

Noise in AWGN channels affects signal bits and is effectively described by Bit-Error Rates (BER): the ratio of incorrectly received bits to the total number of received bits. Each error is statistically independent, and the noise is random and incoherent. In statistical models, the AWGN channel is called a memoryless channel.

Rayleigh channels are referred to as channels with memory in statistical models. Noise events are highly correlated, and appears as block errors rather than bit errors: lengthy sequences of transmitted bits are corrupted or misheard. BER is not usually an appropriate concept.

In research dealing with channel coding that models its effectiveness in overcoming noise, the AWGN channel is treated much more commonly than the Rayleigh channel, for the simple reason that the statistical processes are linear and much easier to model; studying the Rayleigh channel generally requires simulations (typically using Markov chains and Monte Carlo techniques) that can be lengthy and complicated, in order to achieve adequate statistical results.

The maritime RF channel certainly will experience AWGN noise, but our discussion so far should indicates that it is not predominantly an AWGN channel, and we see no indication that it should experience unusual levels of BER compared to other UHF environments that will not be dealt with by protocols schemes that attack the maritime problem of signal fading caused by waves. In other words, the maritime RF channel is primarily a fading, Rayleigh channel, and any protocol scheme must address the problem directly.

Theoretically, there is a paucity of directly relevant analytical results to guide us in detail as we make choices in our low-level protocols. Practically, we can nevertheless arrive at an effective combination of approaches whose design details can then be determined by selective simulation in Phase II.

4.) Protocol Strategy

We can now make a more precise statement of the goal of our research:

> *Look at low-level protocol strategies that will have the greatest utility in mitigating the predominantly fading maritime-communications channel.*

By "low-level protocols", we largely mean the Physical Layer, and to some extent the Data-Link Layer (the distinctions are somewhat vague, and not very important), to use the OSI Reference Model nomenclature; by analogy with a familiar Internet connection over an Ethernet network, these layers are in the less-familiar layers of the network protocol stack below the TCP/IP protocols.

Conceptually, we are considering the transformations of an arbitrary bitstream of data into a frequency-modulated UHF signal. In Figure 2 we break this into two processes: modulation and demodulation (how the RF carrier is altered in time to carry digital information), and channel-coding and decoding (transformations of the bitstream before modulation). To the physical layer, the information that is represented by the bitstream (which itself likely contains packets of data made from a higher-level bitstream, and so on up the protocol stack) is irrelevant.

The following are the relevant tradeoff areas that we consider in the remainder of this report:

- UUV System Configuration
- Satellite Access Methods
- Physical Layer: Modulation
- Physical Layer: Coding
- Higher-Level Protocol Directions & Recommendations

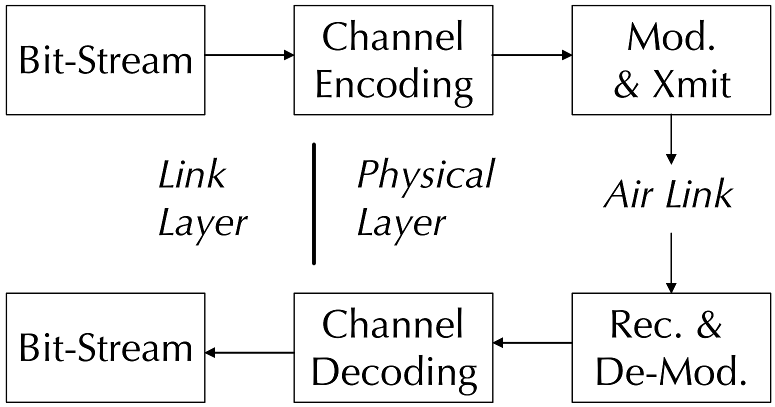


Figure 2: Bitstream Transformation Processes

5.) UUV System Configuration

There are two physical system techniques, beyond the scope of protocols, that can mitigate the problem with waves. One is the simple expedient of mounting (or deploying) the RF antenna on the highest mast that satisfies other system constraints. Since the problem of seeing past waves to low-elevation satellites is a geometric one, even a small increase in the height of the antenna above the sea surface can lead to noticeable improvements in the RF channel. This possibly simple step would improve the reliability of both transmit and receive channels.

The second is the most effective method for reducing the effects of multipath interference in the receive channel: antenna diversity. Conceptually, one uses two (or more) receive antennas, each with its own FM demodulator and possibly channel decoder. The system then switches between whichever receiver has the stronger signal.

Obviously, this is not a simple recommendation to implement. It requires significantly more RF hardware and additional electronics to combine the receive channels. It also require some spatial separation between the receive antennas: to have any effect, they need to be separated by about a wavelength (1 meter). However, the receive system improves significantly in resistance to multipath fading effects.

In addition, we note that transmitted power is always a delicately balanced system consideration. As it concerns RF communications, many problems can be overcome and noise issues dispensed with if transmit signal powers are simply made high enough. Unfortunately, system considerations always limit available power.

- ☑ Raise the receive and transmit antenna on the highest mast that is consistent with other system requirements.
- ☑ In the receive channel, implement antenna diversity to reduce the effects of fading from multipath interference.

☑ Use the most power in the RF transmitter that is consistent with other system requirements.

6.) Satellite Access Methods

There are generally four methods in wireless communication by which a user connects to the network:

- **Single-User Channels**: One user at a time is allowed to use the satellite and all of its available bandwidth resources.
- **Frequency Division, Multiple Access (FDMA)**: The satellite's available frequency spectrum is divided into several client channels, and individual clients are restricted to using the bandwidth of one of the channels.
- **Time Division, Multiple Access (TDMA)**: A communication channel with the satellite is conceptually partitioned into periodic frames, and each frame is further divided into client time slices, thus creating virtual user channels. Each client is restricted to receiving and transmitting during its allocated time slices.
- **Code Division, Multiple Access (CDMA)**: Each client spreads its signal across an available frequency channel ("spread-spectrum radio") by convoluting the transmit signal with another function unique to each user. In effect, the signal from several users are superposed in a way that the receiver can separate again into individual users' signals.

All modern wireless networks now implement some form of FDMA, dividing spectrum in channels that are then used in different ways (e.g., individual cells in a cellular-phone system operate at different frequencies so that signals remain distinct, and individual phones change operating frequencies during handoff as they move from cell to cell). How the available spectrum is divided is determined by bandwidth requirements for individual users according to the type of transmission (voice or data) and the data rates. FDMA is used by all SATCOM satellites.

TDMA and CDMA techniques are used in addition to FDMA (which is so common that it is rarely discussed as such); both are commonly used in modern cellular phone networks. TDMA methods have been in use longer, but are still practical and effective (and the newest GSM standard for cellular networks is a type of TDMA).

CDMA is a newer technology that has been successful in cellular networks. However, if it is to work reliability, it requires dynamic signal balancing. That is, each of several users on a channel must transmit with a power that matches the power of other users' signals at the receiver (the cell tower), or else the signals are difficult to separate. This requires that the network server (the cell tower) continually direct individual clients to adjust their transmit power. We believe that implementing reliable, dynamic signal balancing over a satellite network serving a maritime environment would be technically challenging at this time, and of no appreciable benefit.

The TDMA method is the best choice for the maritime user. TDMA is compatible with the existing SATCOM network (although reports seem to indicate that single-user channels are by far more commonly used), the technology is mature, and it provides adequate multi-user capabilities.

The existing SATCOM TDMA specifications (both the 5-kHz and 25-kHz waveforms) could practically be used by UUVs, although systems considerations may prefer one over the other. However, note that both specifications were written to support voice as well as data communications, and neither is optimized for data-transmission alone. Thus, if a new satellite network were to be used, a new TDMA specification would likely be desirable.

☑ Use TDMA as the multi-user access method with a maritime satellite network, either the existing MIL-STD-188-182 or MIL-STD-188-183 specifications with the current SATCOM network, or a new, data-specific specification with a new satellite network.

7.) Physical Layer: Modulation

a.) Power and BER vs. Eb/N0

A frequently used tool in evaluating channel coding schemes is the graph of BER vs signal-to-noise ratio (SNR) in the signal. Such curves illustrate the fundamental notion that as the SNR decreases, it becomes more difficult for the radio demodulator and decoder to recognize signal bits without error; bit errors are caused by the inability of the demodulator/decoder to distinguish two (or more) states in the received signal, creating the error. Detection of a signal bit is a statistical process, and the probability of correct detection depends on the amount of noise imposed on the signal.

More specifically, the SNR is generally scaled to the ratio of (energy per bit)/(noise per unit bandwidth), or E_b/N_0. An example of such a curve is shown in Figure 3, comparing two frequency modulation techniques.

The specifics of the curves are not relevant, but they illustrate the universal nature of all such curves. At low SNR and high BER, correct detection becomes impossible and the curves flatten to an asymptote. As SNR increases, the noise effects drop off, often sharply.

Such curves are also used in connection with determining link budgets and margins. If it can be determined, say, that an RF channel experiences a BER of 10^{-4}, and QPSK modulation is being used, then the power in the transmitter or gain in the receiver must produce a SNR of greater than 8.2 dB. Conversely, if a system is expected to have a SNR of 5 dB, then the BER will be at least 10^{-2}.

There are two important things to note about these curves. First, the BER is only associated with the statistical detection of the modulation or coding technique for which it is computed; the channel may well have additional noise that must be dealt with. Second, this type of curve has limited utility characterizing techniques that are designed for fading channels, since the noise statistics of the channel are non-Gaussian.

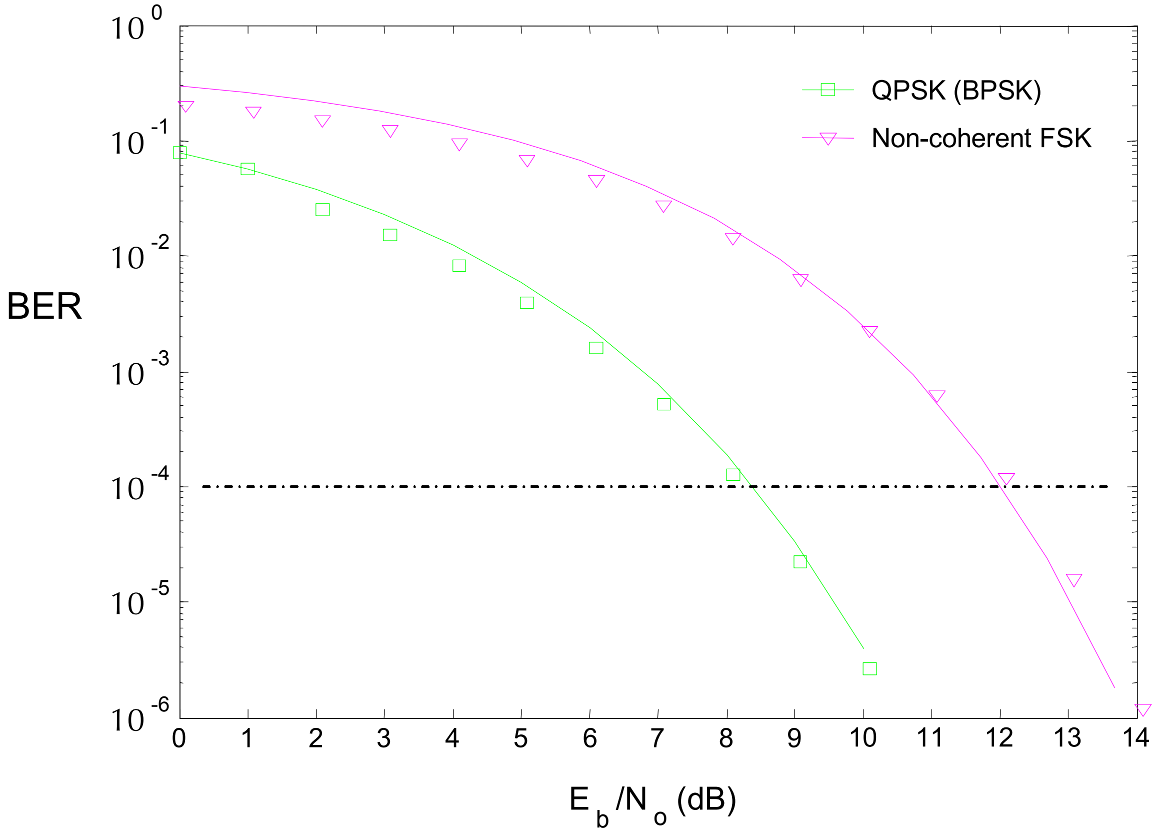
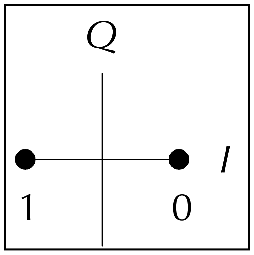


Figure 3: BER vs Signal-to-Noise Ratio

b.) Modulation Techniques

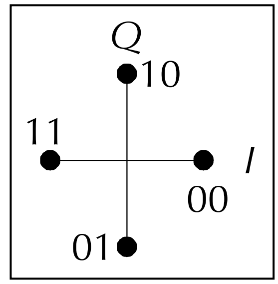
There is a large number of frequency modulation techniques in use for RF communications; the main types, each of which has innumerable variations, are:

- Frequency-Shift Keying (FSK)
- Binary Phase-Shift Keying (BPSK)
- Quadrature Phase Shift Keying (QPSK
- M-ary Phase Shift Keying
- Quadrature-Amplitude Modulation (QAM)
- Trellis-Coded Modulation (TCM)



FSK is the simplest modulation scheme conceptually, and the simplest to implement: as the bitstream switches between "0" and "1", the signal switches between one of two distinct frequencies. BPSK is similar, but instead of switching frequencies to represent "0" and "1", BPSK shifts the phase of the carrier by 180°. This is illustrated in the phase diagram at right: "I" is the in-phase component of the signal, "Q" is the quadrature component. In BPSK, the carrier at any moment is in a state that represents one of the two symbols, "0" or "1".

With QPSK (phase diagram at right), four possible phase states of the carrier are used and its phase at any moment can represent any two bit combination of data from the bitstream: "00", "01", "10", and "11". QPSK requires a more sophisticated implementation than BPSK, but is commonly used.



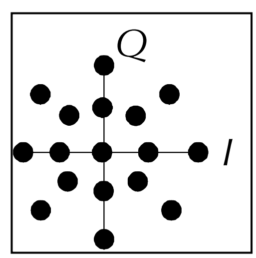
Note that, for each transition in the phase state of the carrier, QPSK can represent twice the data bits that BPSK can represent. Each state transition is called one symbol, or baud, hence the baud rate is the number of transitions per second in the carrier. In BPSK, each symbol has one bit ("0" or "1"), so the bit rate equals the baud rate; in QPSK, each symbol represents two bits, so the bit rate is twice the baud rate. (The confusion between baud rates and bit rate arose when only binary modulation techniques were used and the two were the same.)

For a given baud rate and bandwidth, then, QPSK can transmit twice the digital information that BPSK does, but it comes at a cost. The four states of QPSK modulation are more difficult to distinguish than the two states of BPSK modulation, so more power in the transmit signal is required to overcome the increased effects of statistical noise. QPSK can double the data transmission rate, but twice the E_b/N_0 is needed to achieve the same BER.

The approach of using multiple phases per symbol can be extended to what is called "*M*-ary" Phase Shift Keying; commonly 8-ary and 16-ary are used, in which there are 3 bits and 4 bits per symbol. As with QPSK vs. BPSK, there is an increased cost in transmitted power to maintain a desired BER.

In all of these schemes, the amplitude of the modulated signal remains constant while the phase is shifted. It is possible to increase the number of bits per symbol still further by varying both amplitude and phase, as illustrated in the phase diagram. This is known as Quadrature-Amplitude Modulation, and modern schemes will use as many as 128 bits per symbol.

More sophisticated still than QAM, Trellis-Coded Modulation is a hybrid approach that uses QAM techniques but carefully chooses which symbol will be represented by which phase/amplitude combination in such a way that the receiving demodulator/decoder will have the greatest liklihood of distinguishing a transmitted symbol from symbols that differ in only one or two bits. This adds some level of error recovery, reducing the BER without an increase in power, but with a substantial increase in the sophistication required of the decoder. In effect, the modulation scheme is performing some channel coding (which we discuss next).



Historically, each of these evolutionary steps in modulation has been reflected in the design and speed of modems used over analog phone lines. Early modems, with speeds up to 1200 bps used FSK modulation. QPSK was introduced with 2400 bps modems (i.e., also 1200 baud modems) and also used with 9600 bps (4800 baud) modems. All higher rate modems (28.8 bps and beyond) use some form of TCM, which was introduced as recently as the early 90's.

Digital communication has benefited greatly from these newer, higher density modulation schemes that are able to squeeze more and more data through limited communications channels. They are the result of the realization that information theory (i.e., the "Shannon Limit") limits the information-rate/bandwidth in a communications channel, but not the number of bits/second. However, note that the bit rate can only be increased with a compensating increase in carrier power that can offset the increase in BER that comes from packing more bits into each symbol.

Higher-order modulation schemes are thus very useful in bandwidth limited applications where power is not a severaly limited resource. However, in our application, power and noise are the dominant considerations, while bandwidth limitations are not. We believe that QPSK is the best tradeoff between bandwidth use and power requirements.

We recommend that QPSK modulation be used for its simplicity, its relative efficiency, and its familiarity. QPSK is compatible with existing SATCOM systems (depending on the choice of channel and data rate); we would also recommend it for a new satellite data network for maritime applications. Its familiarity means that it is readily available with off-the-shelf hardware, or can be robustly implemented from standard components in a newly designed implementation.

☑ Use QPSK modulation of the UHF carrier.

8.) Physical Layer: Coding

a.) Channel Coding and Forward Error Correction

Channel coding describes the process by which a logical bit stream is transformed into a modulated signal suitable for transmission of the desired information. Even with a simple modulation scheme like BPSK, it is virtually never the case that a "0" or "1" in the bitstream is directly represented in the modulated signal; indeed, it is likely that the number of symbols in the modulated signal will be greater than the number of bits in the original bitstream. There are many reasons for channel coding, and it is common for several transformations to be applied to the bitstream before modulation.

One form of coding is performed immediately prior to modulation and is designed to aid in clock recovery: providing information in the signal so that the receiver can detect the data rate and synchronize its clock so that bits are correctly recognized. Manchester is one scheme, used in lower speed Ethernet networks, that incorporates timing information into the signal, but at a cost in efficiency.

More common for high-speed networks and wireless communication is a method known as Non-Return-to-Zero-Inverted (NRZI). It is a differential technique, in which a "0" in the bitstream causes a transition in the output (sometimes called "mark"), whereas a "1" causes no transition ("space"). NRZI has the advantage that it contains the same information even if the signal is inverted, which provides some robustness in signal detection.

One potential problem with NRZI arises if a long string of "1" symbols are encoded, which would lead to lengthy time segments in the signal without transitions. The lack of transitions tends to cause loss of clock synchronization in the receiver and a bias in the demodulator towards baseband (DC) levels. This problem has been solved in several ways that ensure frequent transitions in the output; examples include bit stuffing (no string of, say, five "1" symbols are allowed without the insertion of a "0", which will be removed by the receiver), bit-pattern translation (e.g., the "4B/5B translation of fast Ethernet, in which each string of 4 bits is replaced by another string of 5 bits that guarantee at least two transitions), or polynomial scrambling (the bitstream is convoluted in a way that tends to disperse transitions evenly across the signal). Because of its simplicity and lower overheard, we recommend using a simple form of bit stuffing, if necessary, depending on SATCOM compatibility. (MIL-STD-188-182 and -183 do not say whether they specify any type of NRZI encoding; establishing for certain that it is not used is left for a Phase-II activity.)

☑ Use NRZI encoding prior to modulation of the carrier, with bit stuffing for DC load balancing, where compatible with SATCOM requirements.

For our purposes, the more important reason to give careful consideration to channel coding is for Forward Error Correction (FEC). FEC encompasses any number of techniques that transform the bitstream by adding information to it. The extra information allows the receiver to correct bits in the signal that have been corrupted during transmission over the air link. FEC methods add redundancy by adding bits in various ways to the signal, at a cost of increased overhead in the transmission, which thereby reduces effective bandwidth.

The tradeoff issues in selecting FEC methods balance the benefits of increased transmission reliability and the reduced need for retransmission against the costs of bandwidth lost to overhead bits and complexity in implementing the FEC. As noise in the communications channel increases, FEC becomes more of a practical necessity.

Note that FEC is decidedly not the same as error detection, which is done, e.g., with CRC bits embedded in the bitstream by higher level protocols. FEC cannot replace error detection: FEC attempts to deliver a possibly corrupt, received bitstream that is as accurate a representation of the transmitted bitstream as possible; however, error detection remains necessary to validate the accuracy of the received bits and indicate the presence of any remaining errors in the data. Note, too, that not all forms of channel coding (e.g,. NRZI encoding) are error-correcting.

In what follows we will look at several FEC techniques in light of the peculiar requirements of the maritime client, in order to arrive at a balanced solution for reliable communications.

b.) FEC techniques & Coding Gain

FEC methods are divided into two broad categories:

- **Block Codes**: These operate on blocks of bits in the data stream, adding error-correcting bits to the output bitstream. The number of bits in the original bitstream that can be corrected depends on the number of overhead bits added. Block codes are generally better at correcting block errors. Commonly used block codes are:

  - Hamming Codes
  - BCH (Bose-Chaudhuri-Hocquenghem) Codes
  - RS (Reed-Solomon) Codes ☑
- **Convolutional Codes**: These operate on the bitstream in a sequential manner, using shift registers to convolute the stream with itself. Convolutional codes are generally better on isolated bit errors. Examples:
  - Convolutional Codes with Viterbi Decoding ☑
  - Turbo Codes
  - Turbo Product Codes

Block Codes have existed since at least the early 1960s, although they were not used for FEC in communications channels until years later. Convolutional codes are more recent, with Turbo Codes and Turbo Product Codes appearing only in the late 1990s.

The error reduction possible from using a specific method of FEC is often described in terms of coding gain. Referring to Figure 3, imagine that the upper curve is BER vs. SNR for an uncoded data transmission, while the lower curve represents the improved curve due to applying some form of FEC. Reading across as some fixed value of BER (say, BER = 10^{-4}, the dotted line), the SNR of the lower curve is some 82. dB, while SNR for the upper curve is 12.2 dB; this would mean that *with* FEC, a transmitted BER of 10^{-4} can be achieved with a SNR ratio 4 dB smaller than without FEC. The difference, 4 dB in this example, is referred to as the coding gain at 10^{-4} BER. Note that coding gain is always specified at a reference value of BER, commonly at BER = 10^{-5}; the value of the coding gain changes for different values of BER. All commonly used FEC methods exhibit coding gains of about 2 to 4 dB at 10^{-5} BER. (More specific values are quoted in the literature, but the results depend on the details of the noise statistics of a particular communications channel. For our present purpose, these numbers are adequate.)

Turbo Codes and Turbo Product Codes are the most recent developments in creating high-performance FEC codes, developed in the late 1990s. Both rely on methods that use parallel convolutional type encoding, and iterative decoding algorithms with feedback (hence the name "Turbo) in the receiver. The decoding process is complicated, relatively time consuming, and requires substantial computing power to be practical. Standardized, off-the-shelf implementations are starting to appear, although there are still patent issues surrounding Turbo Codes (but not Turbo Product Codes). When Turbo Codes were introduced, they garnered much attention because their performance approached theoretical limits closer than any existing coding scheme; the more recent Turbo Product Codes have a similar performance. However, the coding-gain differential between Turbo Codes and combinations of earlier schemes is small enough compared that we will not recommend using them at this time.

Block codes and convolutional codes have different strengths, and it is possible to combine their operations serially to reduce channel noise further than either can by itself. This is called Concatenated Coding, and has actually been common practice for some two decades. Most notably, the Compact Disc standard developed by Philips and Sony uses concatenated coding that combines Reed-Solomon encoding with Convolutional encoding and Viterbi decoding with a technique called interleaving. We will recommend a concatenated scheme, after discussing each component in the next few sections.

c.) Reed-Solomon Outer Code

Reed-Solomon coding, being a block code, operates on segments of the bitstream, leaving each segment intact but adding a number of bits to the end of the segment that can be used for correcting errors. The theoretical foundation for Reed-Solomon coding is based on Galois-group theory. A particular implementation derives from a specific group that is defined by a generating polynomial; the bits added to a segment make the total block into a divisor polynomial of the generating polynomial.

Decoding Reed-Solomon codes is deterministic, but done at the block level, i.e., received bits are stored until a complete block is received, identifiable errors are corrected, and then the original bitstream segment is output. Depending on the block size there can be short lag times involved. Also, the encoding and decoding processes require complete blocks, so fill bits are occasionally needed.

The theory is complex, the algorithms for encoding and decoding less so, but in practice the algorithms are widely available as firmware for programming FPGA or DSP devices, providing off-the-shelf capability to the radio designer.

Specific Reed-Solomon codes are described as "RS(n, k)", where n is the number of encoded symbols (the total number of bytes in the block), and k is the number of message symbols (the bytes in the original bit stream segment before error-correcting bits are added). The number of error-correcting bytes (not bits!) in each block is $(n - k)$, and as many as $t = (n - k)/2$ symbols (bytes) can be corrected in each block. RS codes operate on symbols, and this gives them their utility in correcting block errors: regardless of where the errors occur in the block, t symbols can be corrected—the errors do not need to be spread across the block.

NASA specifies RS(255,239) or RS(255, 223) for deep-space missions, and these are two commonly options in available firmware. RS(255,239) has 16 error-correcting symbols and can correct 8 message symbols; RS(255, 223) has 32 error-correcting symbols and can correct 16 message symbols. The tradeoff is in the additional overhead from adding error-correcting symbols that reduce the effective throughput of message symbols.

Figure 4 compares BER vs SNR curves for 3 Reed-Solomon implementations (correcting 4, 8 and 16 symbols) with curves for BCH encoding (which we are not considering) and the curve for an unencoded signal. At a BER of 10^{-6}, RS(255,239) gives a coding gain of about 3.5 dB; RS(255,223) gives an additional coding gain of some 1.25 dB, with twice the overhead bytes.

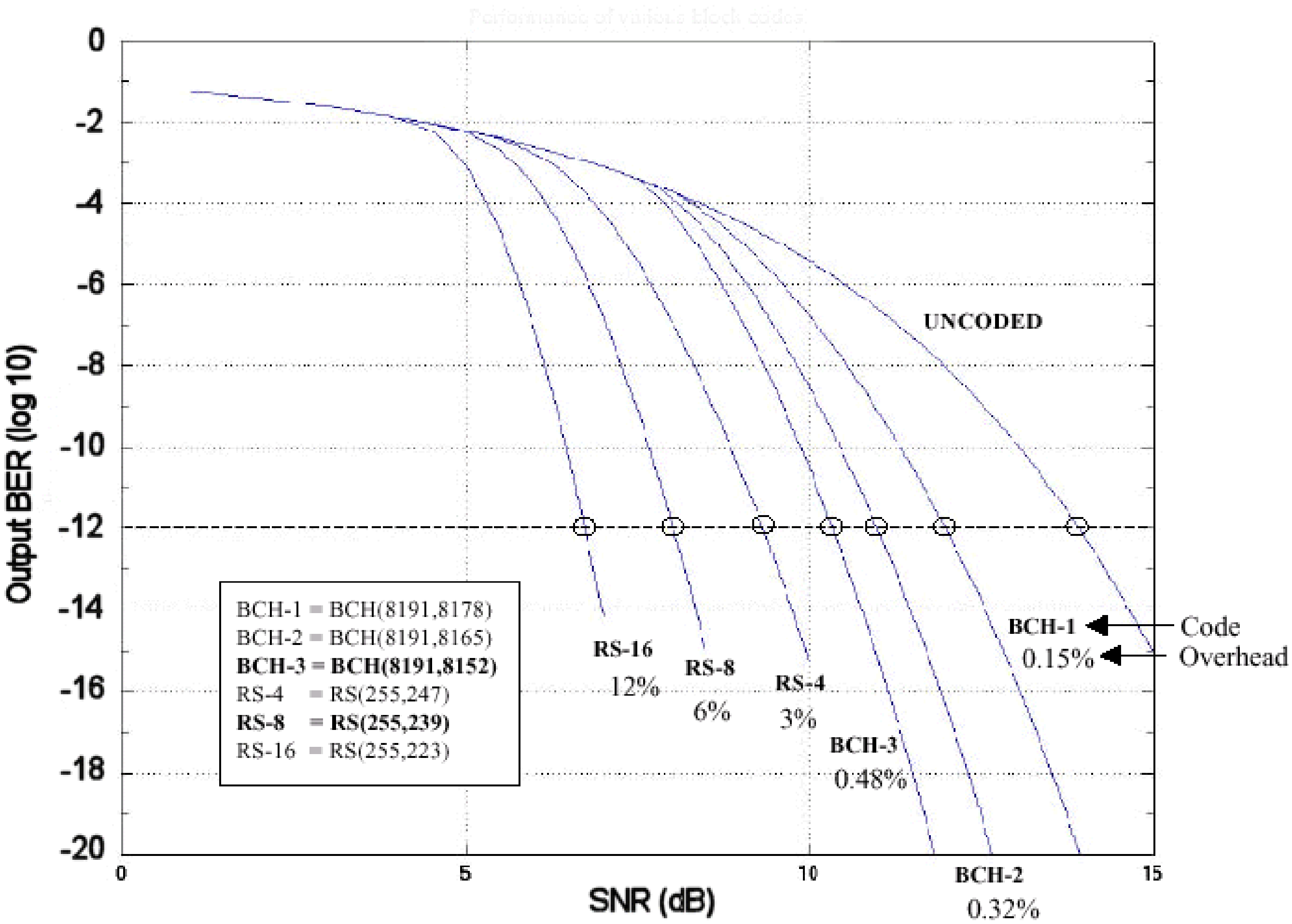


Figure 4: BER vs SNR Curves For Reed-Solomon Implementations

Graph by K. Azadet, IEEE 802.3 High-Speed Study Group, Plenary meeting, Montreal, July 1999

Figure 5 compares the BER of the output of each technique as a function of the input BER. One feature to note is that none of these methods offers any gain at high BER (above about 10^{-2}), which is just a reflection of the fact that the BER is high enough to corrupt more symbols in each block than the RS algorithms are able to correct. However, once into the regime where each is able to correct the bulk of the incoming errors, the output BER declines sharply.

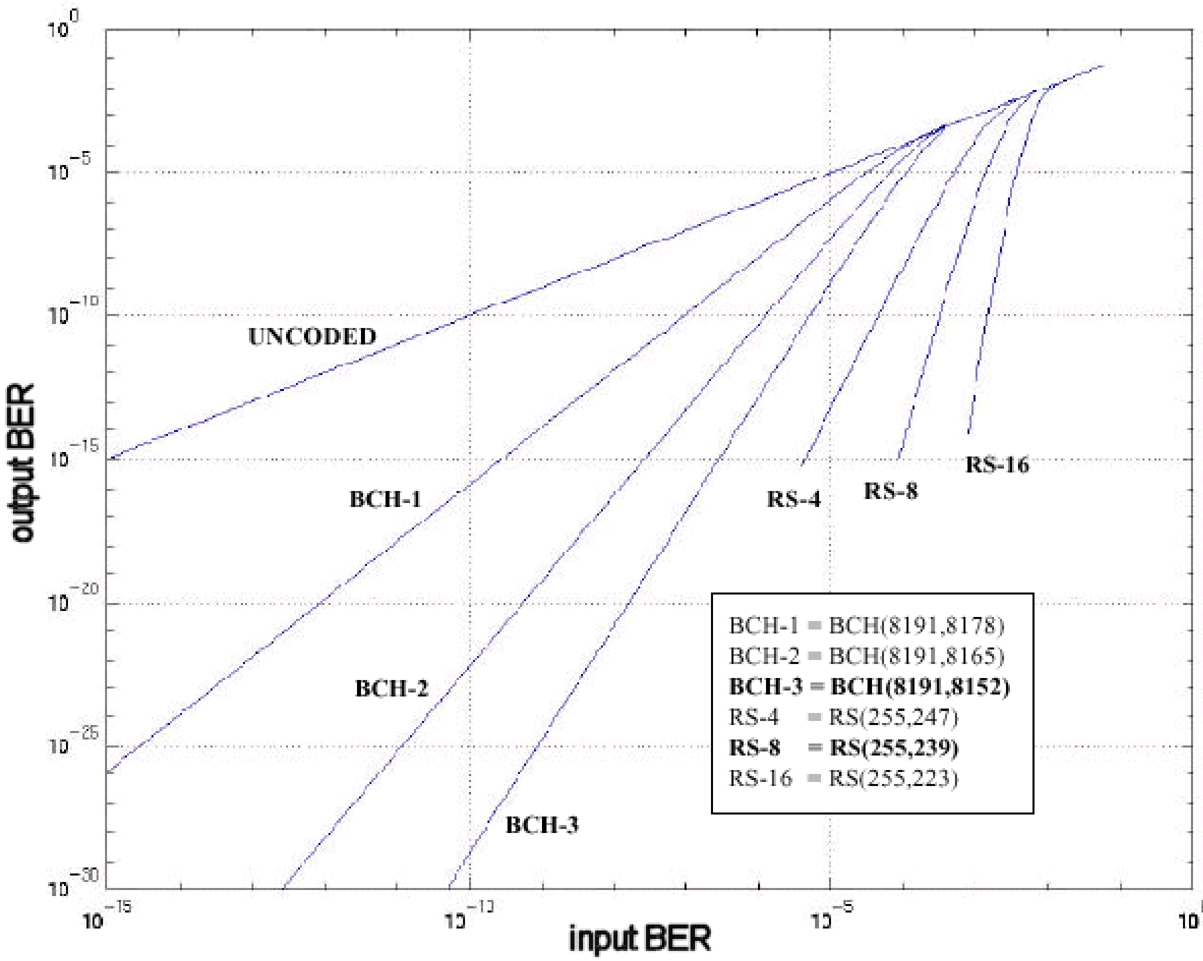


Figure 5: Output BER vs. Input BER For Typical RS Codes

Graph by K. Azadet, IEEE 802.3 High-Speed, Study Group, Plenary meeting, Montreal, July 1999

We believe that RS(255,239) will provide adequate protection against the maritime noise environment with reasonable encoding overhead, and that the additional coding gain from using RS(255,223) is not worth doubling the overhead. However, since both are commonly available in a single, commercial implementation, RS(255,223) could be reserved as an option for unusual cases that merit it, or if detailed analysis in Phase II demonstrates the routine need for the extra margin in coding gain.

☑ As the outer code, use RS(255,239), with RS(255,223) as a design option.

d.) Convolutional Inner Code

Convolutional coding is conceptually simple and generally easy to implement; its practical challenges lie not in the encoding but in the decoding. Figure 6 illustrates the process. The bitstream is clocked into a series of shift registers (labeled "1" through "7" in this example) one bit at a time. At each step, the bits in registers 1, 2, 3, 4, and 7 are XOR'ed and output as bit P1; then the bits in registers 1, 3, 4, 6, and 7 are XOR'ed and output as bit P2.

The output is a bitstream that has twice the number of bits as the original, and each message bit has, in some sense, been spread across 14 of the signal bits. This is the characteristic that makes convolutional coding well suited to correcting random, single-bit errors, and not very robust against block errors. If a single bit in a short sequence of bits is corrupted, nearby bits of the signal still carry information about that bit that can be used to recover its value.

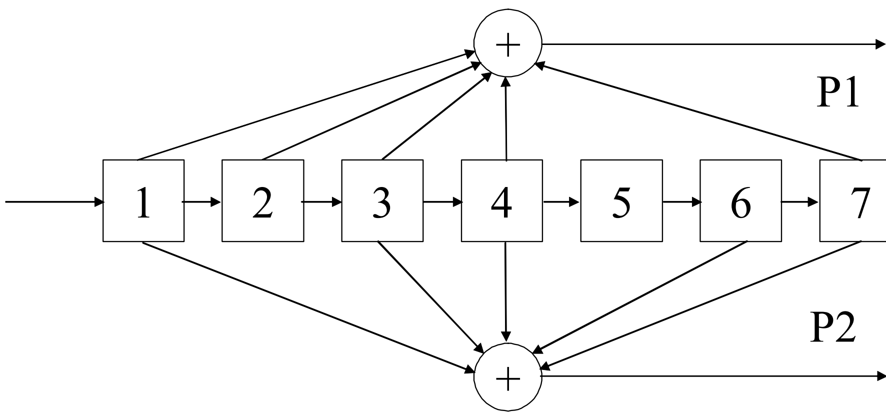


Figure 6: Convolutional Coding

Specific convolutional encoders are described by two parameters: the constraint length, *k*, which is the number of shift registers in the encoder, and the code rate, *r*, which is the ratio of the number of input bits to output bits. The example above is a constraint-length 7, code rate 1/2 encoder.

They are further described by the specific tap positions along the shift registers, referred to as generating polynomial of the convolution (specified as P1 = 1111001 and P2 = 1011011 in our example). Research has provided optimal tap positions for many combinations of constraint length and code rate, and the results are well known. The tap positions shown in the example is optimal for the $k = 7$, $r = 1/2$ encoder. This is also the convolutional encoder specified in both MIL-STD-188-182 and -182; MIL-STD-188-183 (the 25 kHz waveform) also defines a $k = 9$, $r = 3/4$.

The main tradeoff with a convolutional encoder is not in its constraint length, but in its code rate. In this example, which is commonly implemented, $r = 1/2$ means that the effective throughput of the communications channel is cut in half. Surprisingly, it is possible to increase the code rate somewhat without significantly reducing the effectiveness of the encoder, with a technique called puncturing. Briefly, puncturing means simply not transmitting every output bit; in the example above, one puncturing scheme might omit every fourth bit: P1, P2; P1, ; P1, P2; P1, ..., increasing the effective code rate to 3/4. As with generating polynomials, optimal puncturing schemes are well known. As with Reed-Solomon encoders, convolutional encoders are readily available as firmware.

☑ As the inner code, use constraint-length 7, rate 1/2 convolutional encoding.

e.) Viterbi Decoding & Soft Decision

Decoding a convolutionally encoded signal is a non-deterministic process, i.e., there is no closed-form solution to recovering the original message bits. Instead, the decoder implements an iterative algorithm, which we will briefly describe with a simplified example.

Imagine a rate 1/2 convolutional encoder with constraint length of 3, and generating polynomials P1 = 111 and P2 = 101 (which are optimal for this constraint length). The encoder can be represented by a finite-state machine with four states. The state of the encoder is determined by the $k-1$ bits previously shifted into the registers; transitions in the state machine are determined by the bit arriving at each time step at the input. More generally, the number of states is $2^{(k-1)}$, where k is the constraint length.

Figure 7, known as a trellis diagram, graphs the time evolution of the state of the encoder. The circles are the four possible states at each time step, and the arrows represent allowable state transitions, with the state machine starting in state 00. Transitions are labeled as "*m*/*nn*", where *m* is the input bit and *nn* are the output bits of the encoder. For example, if the machine is in state 00, in input bit of 0 causes output bits of 00 and leaves the machine in state 00, whereas an input bit of 1 causes output bits of 11 and a transition to state 10.

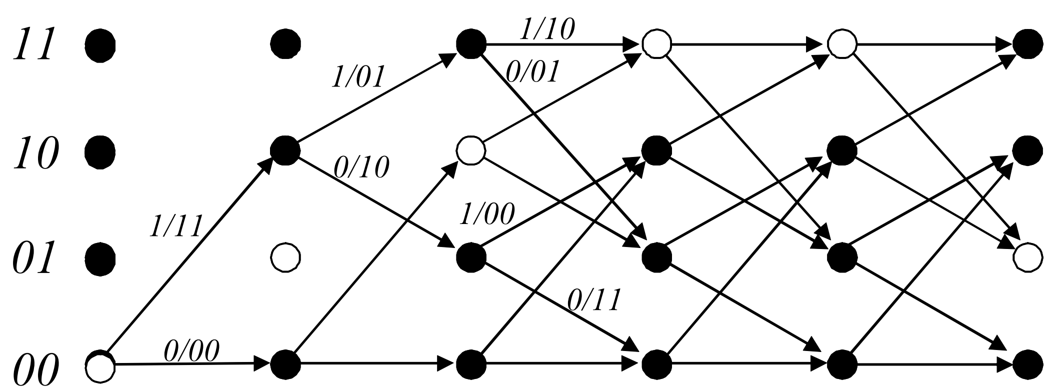


Figure 7: Trellis Diagram

The important point is that, because the convolutional encoder is a finite-state machine *with memory*, its transitions are constrained: from any given state, only a subset of possible states are reachable by allowed state transitions. This is the vital knowledge that allows the decoder to detect and correct bit errors in the received signal, using what is called a Maximum *A Posteriori* (or MAP) algorithm.

The MAP process begins with a segment of received signal, noting its path through the trellis; in the diagram, the path is shown with open signals. In particular, the states that are end points of the path, 00 and 01 above, serve as boundary conditions. Then, all allowable paths through the trellis are constructed and the Hamming distance (the sum of the bits in the difference of the two states) between the each state of the test path and the received signal is calculated at each time step, providing a figure of merit for that test path. The test path with the lowest figure of merit survives the tests and has the highest probability being a correct reconstruction of the original message bits. Then the MAP algorithm repeats by moving the window one time step across the received signal. Iteration between window shifts take place to account for the fact that the boundary states may be erroneous.

The greatest challenge to implementing the MAP-algorithm decoder is the computational complexity involved. With our recommend convolutional encoder of constraint length 7, the number of states in it state-machine trellis is 64, and the number of paths to examine grows geometrically. The contribution of Viterbi was to demonstrate an algorithm that could remove less likely paths from consideration after a small number of steps, thus keeping the required computations manageable. It was the invention of the Viterbi decoder that made high-speed decoding of convolutionally encoded signals possible. For example, all analog-phone line modems that operate at speeds higher than 9600 bps use trellis-coded modulation and Viterbi decoders.

As was the case with our encoder recommendations above, implementations of the Viterbi decoding algorithm are readily available in firmware for use in any new radio design. Since convolutional encoding is an option with UHF/SATCOM, Viterbi decoding will certainly be used in any off-the-shelf, SATCOM-compliant hardware.

There is one further refinement to the process of decoding the convolutional encoding that is worth considering, namely the trade-off between what is called "hard-decision" or "soft-decision" decoding. Simply put, in hard-decision decoding, the decoder is only given the stream of zero and one bits from the radio demodulator to work with, as we described the decoding process in the example above. In soft-decision decoding, signal-strength information is supplied to the decoder along with the demodulated bitstream, in effect weighting each bit to indicate to the decoder whether, say, a "0" is a very definite "0" or only somewhat more likely a "0" than a "1".

Soft-decision decoding significantly improves the performance of the Viterbi decoder, with a coding gain of about 2 dB. Since implementing soft-decision decoding requires hardware support in the radio, it may not be available in off-the-shelf SATCOM compatible radio, but it is a desirable feature if it is available. Likewise, we recommend that any new radio design implement soft-decision decoding.

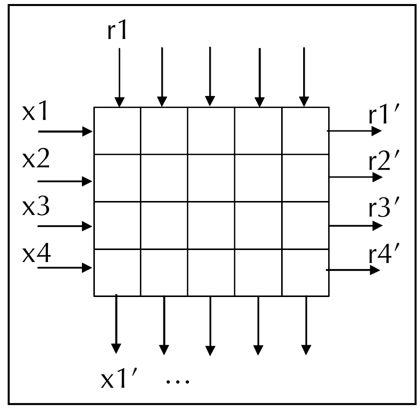
- ☑ Decoding of the inner code is to be done by a Viterbi decoder, implementing soft-decision decoding if possible.

f.) Concatenation & Interleaving

We mentioned previously that Reed-Solomon codes, best at correcting block errors, and convolutional codes, best at correcting bit errors, have complementary strengths and are commonly used together. Combining the two is known as "concatenated encoding." Invariably, the Reed-

Solomon encoding takes place nearest the Link-Layer bitstream and is known as the "outer code", while the convolutional encoding takes place nearest the radio modulator and is known as the "inner code". The two are connected by a vital operation known as interleaving.

The simple idea behind interleaving is illustrated schematically at right. Imagine a rectangular array in the radio transmitter (it need not be square), into which segments of an arriving bit-stream are read in sequentially by rows: x1, x2, x3, etc. When the array is filled, segments are read out sequentially by columns: x1', x2', x3', etc., as the output bitstream.



In the corresponding deinterleaving stage in the receiver, the arriving bitstream is read into the columns and read out by rows, recovering the original order of bits in the original, non-interleaved transmitted signal.

In practice, the interleaving does no have to be so orderly as shown here. Indeed, the SATCOM standards define a random interleaving that operates on blocks of 224 symbols (bytes), mapping them to random positions in the 224-symbol output sequence.

The purpose of using an interleaver between the outer and inner codes may be obvious. The output of the RS encoder is a block of message bits, followed by a block of error-correcting bits. These are dispersed in time before convolutional encoding, and then transmitted. Block errors that might occur during transmission effectively become, after convolutional decoding and deinterleaving, dispersed bit errors that the RS decoder can correct easily. Put another way, interleaving can convert a Rayleigh (fading) channel into an AWGN (memoryless) channel by breaking correlations between statistically dependent bit-error events.

Note that there no error correction associated with interleaving, nor error detection, and there is no associated coding gain. Interleaving is really an error-avoidance technique, but a very important component in the concatenated coding scheme.

Interleaving is particularly important in designing a concatenated code for maritime use by UUVs. Recall two important details about wave characteristics: fading in the communications channel will occur on order one-second time scales and the expected depth of the fading may be 5—7 dB.

A concatenated coding scheme of Reed-Solomon outer code and convolutional inner code with Viterbi decoder and soft-decision decoding will provide a coding gain of about 7 dB with a corresponding reduction in BER from about 10^{-3} to 10^{-9}, but to be most effective they should be combined with an interleaver that matches the time scale of the expected fading events.

Such an interleaver is unusual, and we know of no off-the-shelf implementation with an interleaver depth as large as we recommend. The main reason for this is that the larger the interleaver depth, the longer the lag time in producing the output signal, and that most systems employing these methods over wireless channels must accommodate voice communications, where large-scale interleaving would lead to noticeable and intolerable transmission gaps. Notably, SAT-COM TDMA/DAMA supports voice channels, and its interleaver is on a much smaller scale than would suit our application.

Most applications that use concatenated schemes only use an interleaver to tie together the outer and inner coding, for reasons stated above. However, the specification for CD playback equipment also uses and interleaver following the inner coding prior to signal modulation. Again, the reason is to combat the block errors that are expected to predominate in playback (i.e., scratches and dust on the CD).

We believe it is likely that communications over the maritime wireless channel would also benefit from this extra layer of interleaving. However, characterizing the requirements for pre-modulation interleaving, as well as specifying the detailed requirements on the depth and order of the concatenating interleaving, will both depend on more detailed simulation of the maritime channel and is left as a leading task for Phase II.

- ☑ Concatenate the outer and inner codes with large-scale bit interleaving with a depth that matches the order one-second time scale from the expected fading due to high seas; additionally, consider the utility of a second large-scale interleaver between inner coding and signal modulation.

g.) Metaframing

In our Phase I proposal for this research effort, we introduced the idea of metaframing as an error-reduction concept that we wished to explore, but which we did not develop during our work. Conceptually, metaframing is an approach to implementing TDMA access to the satellite with precisely synchronized signals, so that the metaframing structure can be used as additional information to reconstruct the meaning of partially corrupted, received signals.

We chose to concentrate our efforts in developing a clear picture of the requirements for the concatenated coding scheme that we have presented for two reasons: 1) we felt that more benefit would derive from a careful study and understanding of coding techniques; and 2) it became clear to us early in the task that any metaframing method that we might develop would not be easily compatible with SATCOM systems, and SATCOM interoperability was a high priority.

We believe that this assessment is still correct, and that developing metaframing ideas would not be a suitable task in Phase II, provided that SATCOM interoperability remains a high priority. In the event that there is a move toward a different satellite network system, the metaframing concepts would be worth developing further.

9.) Summary of Tradeoffs & Recommendations

There are many approaches to encoding digital data for transmission through communications channels with varying degrees of noise, more recent and more sophisticated than what we have chosen to recommend. Largely they have resulted from a need to increase bandwidth utilization in high-demand, bandwidth-limited situations, typically with significant need to minimize power requirements.

Instead of adopting the most fashionable techniques, we have chosen a fairly conservative path that will nevertheless meet out needs. Our primary system considerations have been to overcome slow, deep fading in the channel due to high seas, to minimize power requirements and keep computing requirements reasonable, and to retain the greatest amount of SATCOM interoperability possible.

The diagram in Figure 8 presents the protocol stack that we've discussed in this report, showing the hierarchy of implementation from the input of the bitstream from the link-layer on the transmit side, through modulation, transmission, receipt, and back up to output of the bitstream to the link layer on the receive side.

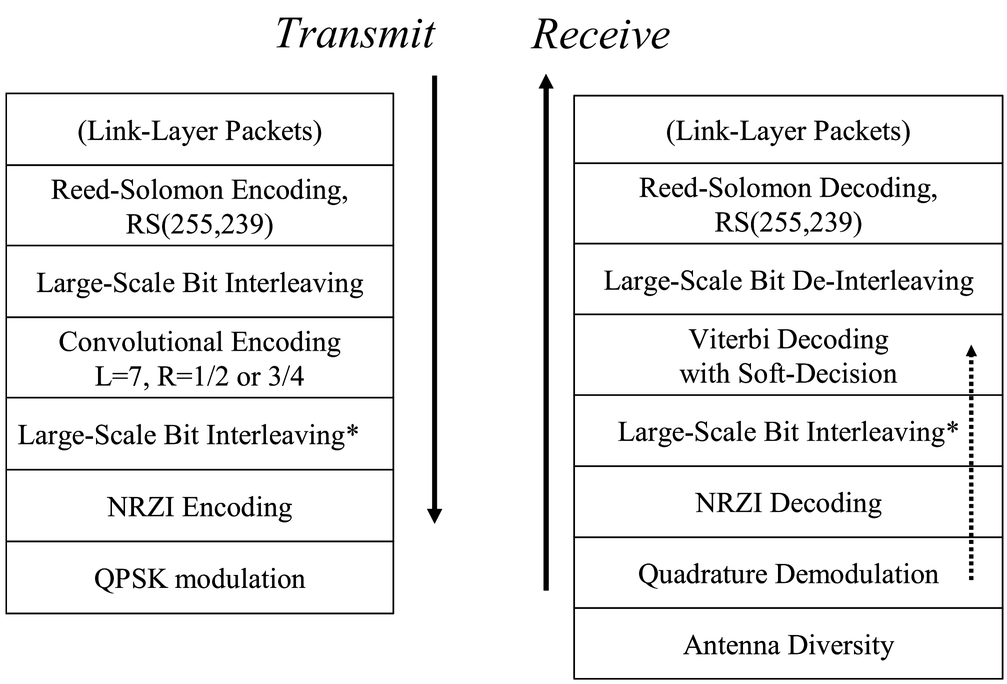


Figure 8: Recommended Protocol Stack

The following list reiterates the individual recommendations that we've made in previous sections.

- ☑ Use a quadrifilar helix antenna, which is relatively insensitive to the ground plane, to avoid concerns about a variable sea surface.
- ☑ Raise the receive and transmit antenna on the highest mast that is consistent with other system requirements.
- ☑ In the receive channel, implement antenna diversity to reduce the effects of fading from multipath interference.

- ☑ Use the most power in the RF transmitter that is consistent with other system requirements.
- ☑ Use TDMA as the multi-user access method with a maritime satellite network, either the existing MIL-STD-188-182 or MIL-STD-188-183 specifications with the current SATCOM network, or a new, data-specific specification with a new satellite network.
- ☑ Use QPSK modulation of the UHF carrier.
- ☑ Use NRZI encoding prior to modulation of the carrier, with bit stuffing for DC load balancing, where compatible with SATCOM requirements.
- ☑ As the outer code, use RS(255,239), with RS(255,223) as a design option.
- ☑ As the inner code, use constraint-length 7, rate 1/2 convolutional encoding.
- ☑ Decoding of the inner code is to be done by a Viterbi decoder, implementing soft-decision decoding if possible.
- ☑ Concatenate the outer and inner codes with large-scale bit interleaving with a depth that matches the order one-second time scale from the expected fading due to high seas; additionally, consider the utility of a second large-scale interleaver between inner coding and signal modulation.

10.) Higher-Level Protocol Considerations

This Phase I research effort has concentrated solely on the Physical Layer, and what might be seen as the interstitial layer of coding protocols between the Link Layer and the Physical Layer, the bottom half of the Link Layer. We've stopped short of considering levels in the protocol stack that begin to operate at the level of data packets, addressing, and channel signaling.

Our reason was simple: the main problem we were trying to solve in this effort was to provide robust wireless communications by mitigating the effects of signal fading in the maritime environment that might result from high seas. Overcoming this noise is a fundamental problem that needed to be attacked at the lowest levels of the protocol stack, and we concentrated out efforts there.

From the start we made a tradeoff decision that Forward Error Correction was significantly more desirable than Automatic Request for Retransmission (known as "ARQ"). ARQ is frequently used in networking protocols, notably TCP/IP, to provide error-free transmission. The idea is simply that packets of data are transmitted, and if some are not correctly received, the receiver requests retransmission.

ARQ, as a convenience to protocols in the stack, is sometimes implemented at a low level. While this may be appropriate for transmission over high-speed, low-noise wire lines, our experience with satellite communication is that it is highly undesirable over high-noise, wireless links, particularly where there is a significant transmission delay. Overall, the satellite link is slow enough, and expensive enough, that it is worth the extra transmission overhead to use robust FEC techniques and to correct as many transmitted packets as possible. ARQ will still be necessary, but should be implemented as high in the protocol stack as possible, to allow the greatest flexibility in tailoring its use to changeable situations.

Our work has focused on the Physical Layer, which is responsible for moving packets across the air link, and the Link Layer, getting packets from the ground user to the satellite. The Network Layer, responsible for switching packets through the satellite network, was beyond the scope of our work. Although any Network-Layer protocol could be imbedded in the schemes we envision, such as TCP/IP, ATM, or Frame Relay, they are not necessarily suitable to the needs of satellite communication.

The main consideration for the higher-level protocols is to recognize the delays inherent in satellite communications. Although delay times with ground-to-ground transmission through geosynchronous satellites may be only half a second (or shorter, with lower orbiting satellites), as satellite networks grow and ultimately incorporate nodes outside Earth's orbit, transmission and switching delays will grow appreciably and must be planned for in a comprehensive, integrated satellite-communications network.

The awareness of the need for delay-tolerant network is just emerging. Research is in the early stages and ongoing, but the groundwork has been laid by a series of "Recommendations for Space Data Systems Standards" by the Consultative Committee for Space Data Systems (CCSDS) [reference 2]. That work would be the starting point for further research into the design of higher level protocols for use with satellite networks.

D. Phase 2 Recommendations

We believe that our list of recommendations is a coherent approach to creating a robust, satellite communications link in the maritime environment, and we have considered the tradeoffs with SATCOM interoperability as a high priority. Nevertheless, not all of our recommendations are compatible with SATCOM specifications, nor are they all equally easy to implement: some will be met with available, COTS products, others would require the design of a new radio system for the UUVs.

1.) Hierarchy of Recommendations

We might categorize our list into three categories of implementation:

- **Easy**: implement within UHF SATCOM TDMA/DAMA capability, likely with COTS equipment.
- **Medium**: protocols go beyond SATCOM capabilities, but possibly implemented over dedicated SATCOM channels; may require new radios and may affect UUV system design
- **Hard**: not compatible with existing SATCOM specifications

2.) Easy Recommendations

Within the **Easy** category, we've placed:

- Use a Quadrifilar Helix Antenna: SATCOM is indifferent to the choice of antenna.
- Use SATCOM TDMA/DAMA satellite access: TDMA is still the best access method, and the SATCOM implementation is acceptable.
- Use QPSK modulation: it is available with several SATCOM channels, depending on the chosen data rate.
- Use convolutional inner code: SATCOM specifies an optional constraint length 7, rate 1/2 code that should be used.
- Add additional interleaving and RS(255,239) outer code *if possible.*

Concatenating the Reed-Solomon code, with intermediate interleaving, may be possible with exiting COTS equipment, depending on where the input interface is provided with the radio internals. If some SATCOM-compatible COTS equipment has made it possible to provide a bitstream directly to the convolutional encoder, then the additional encoding could be implemented externally. However, existing radio designs may make that either impractical or impossible, in which case implementing the very desirable concatenated coding scheme would require a new radio design.

3.) Medium Recommendations

In the **Medium** category:

Implement antenna diversity

- Implement custom-designed radio with full concatenated channel coding:
  - Reed-Solomon outer code: RS(255, 239)
  - Interleaving (depth to be determined)
  - Convolutional inner code: length 7, rate 1/2 or 3/4
  - Pre-modulation interleaving
  - Viterbi decoder with soft-decoding
- Possibly redesign TDMA specifications (over dedicated SATCOM channel)

We note that there may be additional reasons why COTS equipment may not be desirable from a system viewpoint, even if it is possible to create a concatenated code with existing equipment. The SATCOM TDMA/DAMA is a set of complicated specifications that satisfy many potential users in a wide array of operational scenarios. Most of those scenarios involve land-based users, radio systems with directional antennas, and voice communication. Although radio equipment can be SATCOM compliant without implementing all of the specifications, it is unlikely that COTS equipment will be optimized for the maritime user. Designing a new radio for UUVs may well be cost effective and desirable from system considerations.

4.) Hard Recommendation

In the **Hard** category, we place only one recommendation which is not actually in the list, but is implicit in our report:

- Build a new, non-geostationary satellite system that will give significantly better coverage over the oceans and ease the RF communication channel's susceptibility to fading because of high seas and small look angles.

Admittedly, this is a challenging recommendation. We've tried to demonstrate in this report that many of the difficulties faced by the desire for reliable, robust satellite communications for UUVs anywhere in the world's oceans can be mitigated with careful protocol design, there is really no other solution to the geometrical difficulties of communication between a high-latitude maritime client and a geostationary satellite network.

E. Bibliography

1.) Notable Sources

[1] Elwyn Berlekamp, "The Performance of Block Codes", *Notices of the AMS*, **49**, number 1, January 2002.
An accessible introduction to block-code concepts and performance.

[2] Consultative Committee for Space Data Systems, various recommendations available through the website <http://www.ccsds.org>, including "Telemetry Channel Coding", *CCSDS 101.0-B-5*; "Packet Telemetry", *CCSDS 102.0-B-5*; "Space Communications Protocol Specification (SCPS)-Network Protocol", *CCSDS 713.0-B-1* (CCSDS Secretariat, NASA, Washington, DC).
The CCSDS has created a comprehensive set of recommendations for Earth-to-Space data communications that could serve as a starting point in designing a new communications protocol for the maritime-satellite communications channel.

[3] Daniel J. Costello, Jr., Joachim Hagenauer, Hideki Imai, and Stephen B. Wicker, "Applications of Error-Control Coding", IEEE Transaction on Information Theory, 44, number 6, pp 25321-2560, October 1998.
An excellent and comprehensive review of forward-error correction techniques, their application and performance.

[4] Department of Defense, "Military Standard: Interoperability Standard for 5-kHz UHF DAMA Terminal Waveform", *MIL-STD-188-182*, 18 September 1992; and "Military Standard: Interoperability Standard for 25-kHz UHF TDMA/DAMA Terminal Waveform", *MIL-STD-188-183*, 18 September 1992.

[5] Julius Goldhirsh and Wolfhard J. Vogel, "Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems: Overview of Experimental and Modeling Results", ASA-98-U-0-021 (The Johns Hopkins University Applied Physics Laboratory) and EERL-98-12A (The University of Texas at Austin Electrical Engineering Research laboratory), December 1998. Available online: <http://www.utexas.edu/research/mopro/index.html>.
An excellent review article, and the only one to address characteristics of the RF channel over water, albeit only at frequencies above 1 Gigahertz.

2.) General Sources

[6] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding", *The Telecommunications and Data Acquisition Progress Report 42-126* (Jet Propulsion Laboratory, Pasadena, CA), August 15, 1996.

[7] Ezio Biglieri, Giuseppe Caire, and Giorgio Taricco, "Coding for the Fading Channel: a Survey", *Signal Processing for Multimedia* (IOS Press, 1999).

[8] A. Chockalingam, Weiping Xu, and Laurence B.. Milstein, "Performance Analysis of a Multichannel Wireless Access Protocol in the Presence of Bursty Packet Losses", *Proceedings IEEE Military Communications Conference MILCOM* (Bedford, MA, October 1998).

[9] V.Y.Y. Chu and P. Sweeney, "Channel Modeling and Error Control Strategies for the LEO Satellite Channel", <http://www.ee.surrey.ac.uk/CCSR/Software/OPNET/chu-sweeney.pdf">.

[10] S. Dolinar and M. Belongie, "Enhanced Decoding for the Galileo Low-Gain Antenna Mission: Viterbi Redecoding with Four Decoder Stages", *The Telecommunications and Data Acquisition Progress Report 42-121* (Jet Propulsion Laboratory, Pasadena, CA), May 15, 1995.

[11] Intel Technical Staff, "Forward Error Correction Encoding and Decoding", *Intel Application Note 108* (Intel Corporation, 1990).

[12] International Telecommunications Union, "Propagation Data Required for the Design of Earth-Space Maritime Mobile Telecommunication Systems", *Recommendation ITU-R P.680-3* (ITU, 1990-1992-1997-1999).

[13] Jay Klein, "FEC Alternatives for BWA Burst Communications", *IEEE 802.16.1pc-00/23* (a contribution to the IEEE 802.16 Broadband Wireless Access Working Group), April 24, 2000.

[14] John Muhonen and Robert C. Durst. "Performance of Transport Protocols over Satellite Communication Links", *Proceedings IEEE Military Communications Conference MILCOM* (Bedford, MA, October 1998).

[15] Lee E. Taylor and Fran Ganaden, "UHF DMA Satellite Communications Network Control Station: Capabilities, Architecture, and Performance", *Proceedings IEEE Military Communications Conference MILCOM* (Bedford, MA, October 1998).

[16] Leif Wilhelmsson and Laurence B. Milstein, "On the Effect of Imperfect Interleaving for the Gilbert-Elliott Channell", *IEEE Transactions on Communications*, **47**, number 5, pp. 681-688, May 1999.

[17] Yongjun Xie and Yuguang Fang, "A General Statistical Channel Model for Mobile Satellite Systems", *IEEE Transactions on Vehicular Technology*, **49**, number 3, pp 744-751, May 2000.

[18] Michele Zorzi and Ramesh R. Rao, "Error Control Strategies for the Wireless Channel", *Proceedings of ICUPC '96* (Cambridge, MA, Sep-Oct 1996).