HF PACKET: WHERE DO WE GO FROM HERE?

Barry D. McLarnon, VE3JF
2696 Regina Street
Ottawa, Ont., Canada K2B 6Y1

Introduction

Packet radio on the HF bands is alive and well, and is steadily gaining in popularity. The HF links which have been established between widespread packet bulletin board systems have become a workhorse in moving error-free traffic beyond the limits of the VHF/UHF packet networks. These links will, of course, never be capable of handling huge volumes of traffic, like megabyte files: the bandwidth simply isn't sufficient. Satellites and expanded UHF/microwave links must be developed to meet these requirements; but it is probably safe to say that HF will always have a role to play in amateur data communications, both as a back-up to these higher-capacity (but more vulnerable to failure) systems, and for extending the network into remote areas where setting up a satellite station may not be feasible.

There is little doubt that HF packet will play an important role in amateur data communications for many years to come. On the other hand, even its most enthusiastic devotees would likely admit that the performance of the present AX.25 HF packet links is often disappointing. At times, they sail along so smoothly that they are reminiscent of VHF links (albeit at a lower data rate). At other times, for reasons which are often unclear to the users, the links bog down with retries or fail completely, in spite of what appears to be adequate propagation to support communications.

The reasons for this erratic performance can be broken down into three main areas:

(1) The unsuitability, in some respects, of the AX.25 protocol itself for the HF environment.

(2) Difficulties in applying the networking concept of multiple-access (channel-sharing) to the HF environment.

(3) Problems with the modulation schemes and reception techniques used to transmit the AX.25 frames over the HF channel.

In contrast to conventional RTTY and AMTOR, AX.25 is based upon a standard (CCITT Recommendation X.25) which was specifically designed for computer-to-computer communications. On the other hand, it was certainly not designed for use on the HF channel; its usual domain is the relatively benign environment of the Public Switched Data Network.

An important aspect of protocol performance is throughput efficiency. AX.25 does not fare too well in this area, due to the large amount of overhead bits (bits other than information bits) contained in every packet. The overhead amounts to 152 bits, of which 112 are callsign information, assuming a point-to-point link without digipeaters. This is not a serious penalty when maximum-length packets (256*8 = 2048 information bits) are transmitted. Unfortunately, on HF channels the probability of receiving a packet without errors tends to fall off rapidly with increasing packet size, and in practice, much shorter packets must normally be used. The overhead then becomes an appreciable fraction of the total packet length, and the throughput suffers accordingly. Another problem with AX.25 is its inability to take good advantage of longer, multiple-frame transmissions, which reduce the overhead due to turnaround time (transmit/receive switching and transmission of ACK packets). The limitation is a result of the lack of a selective repeat capability in the protocol. The structure of the AX.25 protocol also does not lend itself to the use of signal processing techniques (memory ARQ, forward error correction, soft-decision decoding) which allow error-free packets to be built up from several corrupted packets.

These and other aspects of data link protocol design for HF are treated in a longer version of this paper which has been submitted to the ARRL for publication. Space does permit including the discussion here; nor, I suspect, would the prospect of creating Yet Another Packet Protocol (apologies to WA7MBL) be greeted with widespread enthusiasm in the packet community! The remainder of this paper will be confined to discussing performance improvements which are applicable to AX.25 HF links.

The next reason for poor performance of HF packet links is more a function of usage than protocol design. Packet allows the sharing of channels, by virtue of its CSMA (carrier-sense multiple access) capability. This leads to a tendency for users to congregate on a small number of channels, which is not in itself a bad thing; for low-density traffic like keyboard-to--keyboard chitchat, it can make more efficient use of the limited HF spectrum available. Even for transmission of larger amounts of data, such as file transfers, the lower throughput caused by channel sharing may sometimes be acceptable. Unfortunately, rather than degrading gracefully, the throughput tends to rapidly fall to zero as the number of users on the channel increases. The reason for this is that the collision avoidance mechanism is imperfect. Collisions can occur for many reasons; the colliding packet that zaps yours may come from a "hidden" station in your skip zone, or because a fade caused the carrier detect to fail momentarily. Carrier detect circuits need a certain amount of time to respond; you and the other station may have both started to transmit during that response time "window" (random backoff for retries does help prevent repeated collisions of this type). Possibly the other station isn't detecting you because its receiver is mistuned. Whatever the reason, collisions in multiple-access channels are a major impediment to getting any sort of reasonable throughput.

It also should be mentioned at this point that it is often unclear, particularly to newcomers to the mode, how to set their TNC parameters for best performance on HF. Some rules of thumb have emerged, such as keeping packets short (80 characters or less) and sending only one or two frames per transmission, but the optimum parameters will vary widely with conditions. There is considerable latitude here for experienced operators to "fine tune" their parameters as conditions change. The best bet for newcomers is to check with some of the HF packet gurus and find out what has worked best for them. Some aspects of HF propagation, such as the MUF for a given path, are fairly predictable. An interesting open question for HF BBS operators in particular is the extent to which optimum TNC parameters can be predicted and included in their forwarding files, or perhaps even adapted dynamically.

The balance of this paper will deal with the third topic mentioned above, the modems and associated RF systems used in HF AX.25 systems. First we need a bit of background on HF channel characteristics, and how they affect data communications.

The HF Channel and Modem Design

A large part of the problem with HF packet rests with the design of the modem, which was adapted from the Bell 103 standard. Stated simply, this modem is not capable of reliable communications at 300 bps under the variety of conditions encountered on HF channels. Signalling using binary FSK at this rate produces a symbol length (bit duration) of 3.3 ms. Unfortunately, the signal received at the far end of the link does not usually arrive by means of a single ionospheric mode; instead, it is a superposition of several replicas of the transmitted signal which have travelled by different routes (e.g., they may have undergone different numbers of hops and/or reflected from different layers), and consequently arrive at slightly different times. This phenomenon, known as multipath propagation, is virtually always present to some degree; it results in noticeable distortion to voice signals, but its effects can be much more catastrophic for data signals. It leads to a form of self-interference called intersymbol interference, in which a bit can be demodulated in error due to the delayed energy arriving from the previous bit(s). QRM is bad enough on the bands without doing it to yourself:!

The degree of multipath present in a received signal is measured by a parameter called the multipath spread. Its exact definition need not concern us here; the important thing to know is that when the multipath spread increases beyond a few per cent of the symbol time, the bit error rate performance of the FSK system begins to deteriorate rapidly. When the multipath spread exceeds about 10% of symbol time, it becomes the dominant mechanism in controlling the bit error rate. In other words, if you are operating in this region, improving the signal-to-noise ratio at the receiver (by increasing transmitter power or antenna gain, for example) will produce no significant improvement in error rate! Obviously one should attempt to avoid this situation as much as possible.

Measured data on multipath spreads are not easy to come by, but the author has made a number of observations on short-range paths (60 - 1000 km) in connection with HF data system tests, and would estimate the average spread to be in the neighborhood of 1 ms, with values up to 3 or 4 ms not uncommon. Some data on longer-haul paths have been published (Ref. 1). For example, measurements taken over a four-year period on the 6000 km path between Washington, D.C. and London show an average multipath spread of about 1.3 ms. The observed spread was less than 1 ms for only 30% of the time, and it exceeded 3 ms for 5% of the total period. Similar observations on the 9600 km path between Tokyo and London yielded even higher values: an average spread of about 2.4 ms, less than 1 ms for only 5% of the time, and greater than 3 ms for a whopping 19% of the time!

In addition, there are other impairments which increase in severity with decreasing symbol time. One example is Doppler spreading, a phenomenon most prevalent on signals which traverse the aurora1 zones; it results in the well-known "Arctic flutter" effect. Interference from other stations can also be expected to be more severe as one increases the bandwidth in order to accommodate faster signalling rates. The conclusion is clear: whether the path is short or long, a conventional FSK modem running at 300 bps is not going to deliver a useable error rate for a significant proportion of the time that the band is open and providing an adequate signal level. During these times, increasing transmitter power or antenna gain will NOT help. It is nice to be able to run at 300 bps when the channel supports it, but one should be prepared to fall back to a lower rate, say 75 or 100 bps, when it does not. Better yet, the 300 bps modem should be designed to perform just as well as the lower-speed modem!

The nastiness of the HF channel leads one to ask the question, are all nonamateur HF data systems restricted to very low data rates? In fact, the vast majority of them do operate at rates in the neighborhood of 100 bps. To go higher, one not only pays a price in bandwidth, but the price of a suitable modem tends to rise exponentially with bit rate. HF modems are certainly available for rates up to at least 2400 bps, but the main customers are military, to whom cost is seldom an overriding concern! However, the increasing availability and decreasing cost of digital signal processing components should help to bring the techniques involved within reach of the amateur fraternity. The problem of limited bandwidth available in the amateur bands is a much more serious limitation, as the higher data rate signals would occupy the equivalent of a voice channel. The relatively wideband emissions associated with the 1200 bps and higher-speed modems would not be welcome additions to the congested HF bands, and should not really be necessary in the long run as high-speed satellite and terrestrial links and the higher levels of networking to support them become available. At the present time, it might be wiser to focus on the design of a high-performance 300 bps modem than to charge off in a quest for higher data rates.

Parallel Modems

Commercial HF modems operating at medium and high bit rates (i.e., 300 bps and up) can be roughly categorized as parallel or serial. Parallel modems simply multiplex the data stream into several low-speed subchannels which are spaced just far enough apart in frequency to avoid interfering with each other. A prominent (and very costly) example is the modem defined in the military specification MIL-STD-188C, which uses a total of 16 subchannels spaced 170 Hz apart, plus an additional tone used for correction of tuning errors and Doppler shift. Each subchannel has a basic symbol length of 13.3 ms, which normally would correspond to a rate of 75 bps per subchannel; however, the modulation is four-phase PSK, which allows two bits per symbol to be transmitted, for a total data rate of 2*75*16 = 2400 bps. In practice, this modem is often used in an "inband frequency diversity" mode in which the same data bit is transmitted on two or more subchannels simultaneously. This ploy lowers the effective data rate but increases the probability of demodulating the bit correctly (see discussion of diversity reception below).

Application of the parallel modem concept to a 300 bps design would be straightforward. For example, consider the 100 bps rate recommended above for a single FSK channel; for this rate, the recommendations of the CCIR (Ref. 2) for frequency shift and spacing between adjacent channels are 80 to 85 Hz, and 170 Hz, respectively. Thus we might have three parallel 100 bps FSK subchannels with center frequencies of 425, 595, and 765 Hz (these happen to be the first three recommended center frequencies, but many other choices are possible). The three subchannels could be easily constructed from separate FSK "building blocks" similar in design to those used presently by amateurs. The inputs to the three subchannels would be derived from a 1-out-of-3 data multiplexer, and the three outputs summed before being applied to the transmitter. An attractive alternative would be to implement all three subchannels with a single digital signal processor (DSP) chip. A modem using one of these devices would have the considerable advantage of needing no tuning whatsoever - it comes to life with all filters and oscillators perfectly tuned and stays that way! (some analog filtering is needed for anti-aliasing and reconstruction in the analog-to-digital-to-analog conversion processes, but this is relatively noncritical and should never require adjustment). The cost of DSP devices and their development tools has kept them out of amateur applications, but it just a matter of time before they will begin to make their presence felt; they are the wave of the future in low-frequency signal processing.

Serial Modems

The second major category of modem is the serial modem, which simply means that only one signal (normally a sinusoid modulated in frequency or phase) at a time is transmitted. Most telephone-type modems, and al 1 modems presently used by Amateurs, fall into this category. Each signal (symbol) may represent a single bit of information, in which case the modulation technique is called binary (as in the 300 bps and 1200 bps binary FSK modems now used for AX.25 data links); in this case, the signal has two possible states, commonly called mark and space. Most 1200 bps and higher-rate modems produce signals with more than two states and thus carry more than one bit of information per symbol; otherwise, their spectra would not fit within a standard voice channel. Serial modems tend to have much shorter symbol lengths than parallel modems operating at the same bit rate, and therefore they require a more well-conditioned channel in order to avoid intersymbol interference. Most telephone-line modems include an equalizer in order to condition the channel;.it consists of a filter in front of the modem which is designed to flatten the amplitude and time delay response of the channel and thereby reduce the intersymbol interference and other distortion that result in reduced noise margins in the demodulator. Some modems, such as the very common 212A 1200 bps type, used a fixed equalizer design based upon typical telephone channel characteristics. More sophisticated higher-speed modems use an adaptive strategy: at the beginning of the call, a special "training sequence" is transmitted from each end of the circuit which enables the receiving modem to adjust the parameters of its equalizer for minimum distortion of the received data signal.

The principle of adaptive equalization also applies to HF modems, but successful implementation is much more difficult. Since the response of the channel is now time-varying, the equalizer parameters must be frequently updated. This generally means that the training sequence must be periodically reinserted into the data stream to allow re-adaptation. Modems which employ the training sequence technique are generally known as "reference-directed" adaptive modems. Other adaptation algorithms have been developed which do not require special sequences to be transmitted; for example, the equalizer can be adjusted to maximize the demodulator "eye pattern" opening without knowledge of the actual data sequence transmitted. This mode of operation is known as "decision-directed", and is of course more desirable due to the lack of overhead involved. Quite a number of attempts, some dating back to the mid-sixties, have been made to implement adaptive serial HF modems, mainly for the 2400 bps data rate. These designs share a common characteristic: when the channel is reasonably well-behaved (e.g., slow fading), they tend to perform well, sometimes even spectacularly; when the channel gets nasty, however, there always comes a point when the rate of adaptation is not sufficient to keep up with the fluctuations in the channel, and the modem fails equally spectacularly. In the latter case, a parallel modem may still deliver a usable error rate and therefore work over a wider range of conditions. On the other hand, the serial modem may offer higher overall throughput by virtue of better performance during the majority of the time, when the channel is not varying too rapidly. One reason for this better performance is the following: the serial modem transmits a single sinusoid at a time, and thus produces a more or less "constant envelope" signal; contrast this with the parallel modem output, which is a summation of several sinusoids. The parallel modem signal therefore has a higher peak-to-average ratio, and consequent1 y will yield an output signal with lower average power from a typical peak power-limited transmitter (note that the clipping and compression techniques often used to overcome a similar problem with voice signals will not be well tolerated by the data signal!). All things being equal, the serial modem will then have more " sock" to cut through the noise and QRM when these are the primary limitations to communicating on the channel. Considerable effort continues to be expended on adaptive serial designs, and their performance should continue to increase as the state of the art in digital signal processing devices advances.

Although adaptive serial modems are presently difficult and costly to design and build, they should eventually find their way into amateur applications. Making one work well at 300 bps should be considerably less difficult than at 2400 bps and higher. (concentration on the higher data rates was stimulated in large part by a strong military interest in secure digitized speech). One intriguing possibility is the design of an adaptive equalizer to work with the presently-used 200 Hz shift 300 bps modems. Perhaps there is a well-heeled experimenter out there somewhere who needs a challenge!

Variable-rate Moderns

A useful concept which can be applied to both parallel and serial modems is that of the variable-rate modem. The basic idea here is that the channel capacity (the maximum rate at which information can be reliably transmitted) of an HF link with a given bandwidth is not fixed, but time-varying. In order to keep the link reliable, we should attempt to adjust our signalling rate to match the available capacity. In contrast to a fixed-rate modem which is likely to collapse completely in the face of worsening conditions, the variable-rate modem allows the throughput to degrade gracefully. This concept is embodied in the Packet Adaptive Modem described by Rinaldo (Ref. 3). A much more sophisticated design, a parallel modem utilizing DSP techniques to provide six possible rates from 75 to 2400 bps, is described in Ref. 4.

Although simple in concept, the variable-rate modem is tricky to implement, the problem being to develop a suitable algorithm to monitor the system performance and carry out the necessary adaptation automatically. Performance monitoring is not too difficult, but any changes that ensue must be coordinated between the two ends of the link. This calls for a highly robust low speed link piggybacked onto the main data link; such a link is often termed an "order wire". Development of an effective variable-rate modem appears to be a worthwhile objective for the amateur community; in particular, a variable-rate serial adaptive modem would be less difficult to implement than a high-speed fixed-rate serial modem with adaptive equalizer, and it would avoid the peak-power limitations of the parallel modem.

HF Receiver Design

Some aspects of equipment design for packet operation, such as faster turnaround times, are beginning to be addressed by the manufacturers of amateur radio gear. Nevertheless, one suspects that they are less than fully cognizant of the requirements imposed by packet operation, particularly in the area of HF receiver design.

It is probably safe to say that the most important element of the HF receiver used for packet operation is the IF filter. This fact has been clearly demonstrated by Eric Gustafsen (Ref. 5) in his comparative study of HF modems. He also makes the important point that an audio filter, no matter how good, is not an adequate substitute for a suitably narrow IF filter. The crucial difference is that the audio filter will not prevent unwanted signals outside its passband from reaching the AGC detector, resulting in receiver desensitization and cross-modulation on the desired signal. The optimum IF filter bandwidth, of course, depends upon the type of data signal being received. For the 300 bps, 200 Hz shift binary FSK emission in present use, a study done in the early 70's (Ref. 6) indicates the optimum bandwidth should be about 360 Hz. This is a bit narrower than Eric's recommended range of 400 to 500 Hz. One good reason to use a wider than optimum filter at the present time is that available IF filters tend to have severe delay distortion (i.e., nonlinear phase characteristics), and this distortion is worst near the edges of the passband. This is a consequence of designing the filter for maximum possible rolloff rate in the stopband. The variation in delay over the passband of the filter can easily be several milliseconds, which can cause considerable intersymbol interference and consequent higher error rates in the data signal. This distortion could be reduced by means of a suitable equalizer, but hopefully this will become unnecessary as filters with characteristics more suited to data transmission become commonly available.

Another aspect of HF receiver design that is ripe for further study is optimization of the AGC system for packet transmissions. There is little doubt that the slow-release type of AGC time constant used for SSB reception is not very suitable for reception of data, especially when atmospheric noise is severe. However, it is not clear that the faster AGC characteristic commonly used for CW reception is that much better, or whether some other characteristic might be substantially better. In any case, the "optimum" is likely to be dependent on band conditions. Several speakers at an HF communications conference attended by the author in 1985 mentioned the dearth of knowledge concerning optimization of receiver AGC for data transmission. One stated that he had achieved better results by disabling the AGC entirely and carefully setting the RF gain manually.

Diversity Reception

Diversity reception is a technique which has found little application in amateur circles, in spite of the fact that its benefits have been known since the early days of radio. The reader is referred to the article by Nagle (Ref. 7) for a good overview of diversity techniques and their history; here we shall summarize them briefly and focus on their application to packet-type data communications.

Diversity reception might be defined as the processing of alternate versions of the same transmitted information in order to demodulate it more faithfully. The alternate versions may be generated at the transmit end by transmitting redundant information (in which case the technique bears more than a passing resemblance to error-correction coding!), or they may be generated solely at the receive end by sampling the received signal in two (or more) different ways. The first category includes frequency and time diversity, and the second includes space and polarization diversity. The key to success of the technique is that the different versions of the signal have encountered quite different perturbations during their travels through the ionosphere, or, in mathematical parlance, that they are highly uncorrelated.

A straightforward and widely used (in commercial HF links) application of the diversity principle is frequency diversity, in which the same data is transmitted simultaneously on two or more separate frequencies. BY far the most common implementation is dual diversity, with two subcarriers carrying the same data (going to higher orders of duplication than two produces a state of rapidly diminishing returns). The separation required between the subcarrier signals to yield little or no correlation varys considerably with channel conditions; it may be tens of kiloHertz under extremely good, stable conditions, and as little as 100 Hz or so when conditions are very unstable. The minimum separation for decorrelated signal depends upon the multipath spread; a reasonably good "rule of thumb" holds that the separation should be at least one-half of the reciprocal of the multipath spread. For example, when the spread is 1 ms, the separation should be at least 500 Hz. In most cases, the separation used is of the order of 1 kHz; these implementations are known as "inband" frequency diversity, since the data subcarriers are contained within the bandwidth of a single voice channel. Such would probably be the case with any amateur implementations as well. Larger separations would provide better performance, but the technical and regulatory problems become more formidable as well. Even relatively small separations can give worthwhile performance gains, however. One study (Ref. 8) of inband frequency diversity yielded an average improvement in bit error rate of about one order of magnitude over single-channel operation. Such an improvement could result in a dramatic increase in system throughput. In this particular case, the data rate was 75 bps and the subcarrier frequency separation was 1360 Hz.

Returning for a moment to the parallel 300 bps modem proposed above, dual frequency diversity could be added in straightforward fashion by adding three more FSK subcarriers. The fourth subcarrier would carry the same data as the first, and so on. The simplest subcarrier frequency assignment would be to use the next three standard center frequencies, maintaining the 170 Hz spacing. This gives a separation between subcarriers carrying the same data of 510 Hz, which is considerably less than ideal; nevertheless, diversity gain would be available for a good deal of the time, and in particular when the multipath is severe. Other schemes are possible, such as spacing the two groups of subcarriers farther apart and creating a "hole" between them, which could then be occupied by other signals. However, special IF filtering would then be needed in the receiver to remove the unwanted signals in the "hole"; otherwise, these signals could reach the AGC detectcr and desensitize the receiver.

The next major category of diversity operation is time diversity. Here the same data is transmitted two or more times, with a time separation between the transmissions chosen such that the perturbations undergone by the signal are largely uncorrelated from one transmission to the next. Time separations of at least one or two seconds are generally required for this condition to hold. There are a number of practical problems in implementing a scheme of this type. In any case, it can be argued that a system which employs an ARQ protocol already has what amounts to time diversity built into it, and the interval between repetitions of a block of a data will almost certainly be sufficient to guarantee uncorrelated conditions. Furthkrmore, the ARQ system tends to adapt to the channel conditions, since the rate of repeats will be inversely related to the severity of the disturbances. When the channel is good, repeats will be few, and thus it will not suffer the penalty imposed by the fixed amount of redundancy in a simple time diversity scheme.

The next form of diversity reception we shall discuss, space diversity, has some intriguing possibilities. Space diversity involves the simultaneous reception and subsequent demodulation of the signal from two or more physically separated antennas. Once again, the aim is to derive uncorrelated versions of the signal, in this case by demodulating signals which have followed slightly different paths through the ionosphere and hence have undergone different perturbations. Here again, we have a technique which has been widely used in commercial HF applications for many years, and yet has been largely ignored by amateurs. Granted, the physical constraints imposed by many amateur installations may preclude the use of space diversity; nevertheless, the technique is within the grasp of many amateurs. Space diversity has one major advantage over frequency and time diversity: it does not involve the addition of redundant information to the transmitted signal. Since only the receiving set-up is changed, the technique could be applied immediately to existing data transmission techniques as well as future ones, and no regulatory hurdles need be overcome.

And now for the bad news (as usual, there's no free lunch!). In addition to two antennas (the use of dual diversity is assumed hereafter), you will need two receivers and two demodulators.

Duplication of the demodulator portion of the HF modem is not a major problem, but not everyone has two good-quality HF receivers in the shack. For those who do have the requisite receiving equipment, the next major consideration is the antennas. Other than being reasonably similar in gain properties, the primary requirement is that they be spaced far enough apart to yield a worthwhile diversity gain. How far apart is enough? Opinions vary, and actual measured data are scarce. Most textbooks state that the spacing should be nine or ten wavelengths; another (Ref. 9) states the minimum useful spacing to be four wavelengths. The 1985 ARRL Handbook, inexplicably, gives a value of only 3/8 of a wavelength as providing useful gain. The latter value is a bit hard to swallow, although Nagle (Ref. 7) does claim that good results have been obtained with a spacing of around one wavelength. In addition to reducing the potential diversity gain, however, very close spacing may cause problems with the matching of the antennas due to mutual coupling effects. The only thing that is certain is that you cannot have too much spacing, and should try for the maximum that is practical.

An interesting possibility to get around the constraints of small city lots and the need to own two sets of receiving equipment is to make an arrangement with a nearby buddy to use his station as a remote receiving site, and bringing the received audio back to your QTH via a telephone hook-up or low-power UHF link. The major stumbling block here is the need for some type of remote control of receiver tuning, but receivers with this capability are becoming increasingly common these days. Another solution to the antenna spacing problem of space diversity reception is to use a related technique, polarization diversity. A good deal of the fading experienced on the HF bands results from polarization mismatch between the receiving antenna and the signal, which is in turn caused by polarization rotation of the signal as it passes through the ionosphere. Combining the outputs of two co-located antennas, one having vertical polarization, and the other horizontal, can produce a marked decrease in fading and consequent lower error rate.

Having derived a pair of diversity signals by some means, it remains to combine them to produce a single output data stream. To begin with, each signal should be separately demodulated up to, but not including, the point at which a hard decision is made as to which data symbol was transmitted. The corresponding signal is that which is often observed in modem testing as an "eye pattern" (so-called due to the appearance of the signal when observed with an oscilloscope whose timebase is synchronized to the symbol timing of the received data signal). The "eye" signals from the separate diversity paths can be combined by means of one of three basic techniques: linear, selection, or maximal-ratio combining.

The three combining techniques have their theoretical pros and cons, but performance on the HF channel does not always subscribe to the theories! Linear combining is the easiest to implement, as the diversity signals are simply added together (typically with an op amp summing circuit) before being presented to the comparator or whatever circuit produces the binary output data. This simple scheme can work surprisingly well, but is not recommended for situations in which a diversity channel tends to produce a high noise output when the signal fades in that channel. This situation prevails, for example, in FSK demodulators using hard limiters, or in separate-receiver systems in which the receivers each have an independent AGC.

The next step up in complexity is selection combining, in which only one diversity channel is connected to the output decision circuit at any given time. An attempt is made to continuously monitor the signal strength in each diversity path and to rapidly switch to whichever is strongest. In practice, some hysteresis is built into the selection circuit in order to prevent excessive "hunting" back and forth between channels. Selection is clearly suboptimal in that potentially useful contributions to the decision process from the unused channel(s) are thrown away.

The third technique, maximal-ratio combining, in a sense combines the best features of the first two. The diversity signals are summed as in linear combining, but before summation the amplitudes of the signals are adjusted by multiplying them by a weighting factor which is proportional to the signal power in the corresponding channel. This approach makes the best possible use of all of the received signal information, but its theoretical advantages may not always materialize on real-world channels, and the complexity of the circuitry is considerable compared to the other methods. Nevertheless, the design of a maximal-ratio combiner is quite straightforward, and experience has shown that it will generally outperform the other methods by a small margin on the HF bands.

There has been a recent trend towards building more "intelligence" into diversity combining systems. A basic problem with most combiners is that the circuits which measure signal power to provide the basis for selection or maximal-ratio combining are usually "dumb"; that is, they cannot distinguish the desired signal from noise and interference since they simply measure the total energy within a certain passband. This causes errors in selection or weighting to occur which can seriously degrade the performance of the diversity system. The key to overcoming this problem is to make the circuitry which assesses the diversity channels sensitive to certain known attributes of the desired signal. For example, a 100 bps "eye pattern" signal can be fed to a circuit which generates a fixed-length pulse for each zero-crossing of the signal. If the data signal is strong, the pulses will occur at 10 ms intervals, or integral multiples of 10 ms. The frequency spectrum of the pulse train will then tend to have its energy concentrated around 100 Hz and its harmonics. If the signal is dominated by noise and interference, on the other hand, the pulses will be more randomly distributed in time and the spectrum will not exhibit the same concentration of energy. A circuit consisting of two narrow bandpass filters centered on 100 Hz, followed by rectifers, lowpass smoothing filters, and a comparator can be used to distinguish between the two conditions. The output of such a circuit makes a more reliable signal quality assessor than a simple energy detector. Additional details on building intelligence into diversity systems are contained in Ref. 10.

Conclusions

The performance of HF packet systems can and should be improved. Considerable improvement in the performance of the present AX.25 system is possible through the design of better modems, improved HF receivers, and the use of diversity reception techniques. It is hoped that some of the techniques mentioned in this article will help point the way.

References

1. CCIR, Report 203-1, "Path time-delays and shifts caused by multipath propagation on radio circuits", CCIR Green Book, vol. 3, Geneva, 1982.

2. CCIR, Recommendation 436-2, "Arrangement of voice-frequency telegraph channels working at a modulation rate of about 100 bauds over HF radio circuits", CCIR Green Book, vol. 3, Geneva, 1982.

3. Paul Rinaldo, W4RI, "Introducing the Packet Adaptive Modem (PAM)", Second ARRL Amateur Radio Computer Networking Conference, March 19, 1983, pp. 2.71-2.74.

4. J.M. Perl, "Channel coding in a self-optimizing HF modem", Proceedings of the 1984 International Zurich Seminar on Digital Communications", Zurich, March 6-8, 1984, pp. 101-106.

5. Eric Gustafsen, N7CL, "HF Modem Performance Comparisons", Packet Radio Magazine, vol. 2, nos. 1&2, pp. 21-24, January-February 1987.

6. R.T. Bobolin and J.C. Lindenlaub, "Distortion Analysis of Binary FSK", IEEE Trans. Communication Technology, vol. COM-19, pp. 478-486, August 1971.

7. John J. Nagle, K4KJ, "Diversity reception: an answer to high frequency signal fading", Ham Radio, November 1979, pp. 48-55.

8. P.McManamon and R.V. Janc, "An experimental comparison of nondiversity and dual frequency diversity for HF FSK modulation", IEEE Trans. Communication Technology, vol. COM-16, pp. 837-839, December 1968.

9. Keith Henney (ed.), Radio Engineering Handbook, 5th Edition. McGraw-Hill, 1959, p.19-118.

10. T. Gartell and D. Cawsey, "An intelligent diversity system for improved signal quality", Proceedings of the Conference on Communications Equipment and Systems, Birmingham, England, pp. 234-236, 1978.