IEEE 802.11

Wireless Access Method and Physical Layer Specifications

Title: Data Encoding Schemes for Infrared Signaling

Author: Ellen Oschmann
Spectrix Corporation
906 University Place
Evanston, Illinois 60201
Voice: (708) 491-4534
FAX: (708) 467-1094

Infrared optical signaling has demonstrated considerable potential for the realization of high bit rate, untethered data communication channels. Though a number of diverse applications have long existed for such data links, the rapidly growing population of portable computer users and the associated demand for wireless network connectivity have imparted a new urgency to the effort to develop a well defined and widely accepted reference model for this increasingly essential class of peripheral. The present submission, which seeks to address the question of physical layer data encoding, is offered in the hope of stimulating discussion in furtherance of that effort.

The method by which the message bit stream in an infrared signaling system is encoded for transmission is a critically important and fundamental aspect of the design. A well chosen line code is a valuable tool in combating the unpleasant realities of the diffuse optical channel, and in mitigating the inadequacies of commercially available infrared emitting and sensing devices. Manchester [biphase] encoding and Pulse Position Modulation [PPM] are both familiar, useful schemes which warrant serious consideration; it may be, though, that another technique, simple Return-to-Zero [RZ] encoding, is better suited to the demands of high data rate, limited range optical links.

No matter the ultimate choice of encoding scheme, the characteristics of the IR transmitters and receivers constituting the system hardware will demand careful attention. The infrared emitting diodes [IREDs] and photodetectors available today are scarcely the ideal devices for which a system designer might wish, and as their flaws may be as readily meliorated as exacerbated by the coding method they must support, it is as well to take a moment now, before considering the matter of encoding, to become familiar with a few salient aspects of their behavior, and of the medium itself.

IREDs and Photodiodes

As cost and safety issues will in fact rule them out of most portable designs, semiconductor laser diodes will not be considered. IREDs, the remaining option for transmitting sources, are most efficient when operated at lower peak power levels; the finite slope of the forward conductance curve implies an internal ohmic power dissipater which accounts for part of this effect, whilst the remainder is mostly a consequence of carrier saturation at very high forward currents. Not only does efficiency dictate low IRED power levels -- so too does the desire to maximize the longevity of the devices, for high peak currents generally result in a more rapid permanent degradation of optical power output, often the result of the formation of dark line defects in the IRED die, and in a greater likelihood of catastrophic device failure.

While IREDs available today have optical transition times sufficiently short to support operation at 10 Mbps, it is not true that the turn-on and turn-off times are always equal; even when the data sheet promises symmetry of operation, the necessarily high drive current slew rate, easily twenty amperes per microsecond, can thoroughly confound an inadequately designed driver circuit, masking the potential of even the best IREDs. Unequal IRED transition times are a source of bias distortion, a mechanism that can increase the probability of transmission errors.

An oft-repeated but fallacious belief states that, because the PIN photodiodes commonly used as receptors in direct detection IR receivers produce an output current proportional to incident optical power, output electrical power is a function of the square of input optical power, and it is therefore advantageous to signal with the highest peak optical power possible. The truth is that, ideally, the photodiode never transfers any power at all, for it typically looks into a virtual ground [at the input of a transimpedance amplifier], and the "higher peak transmit power is better" argument loses its validity, except when the associated receivers use very basic threshold-sensitive data regenerators -- and even then, performance improvement with increasing transmit power is only linear at best.

As regards the character of the optical signaling medium itself, it is very important to note that it is inherently unipolar. One cannot launch optical pulses of two distinct polarities; an emitted pulse can only ever *add* energy to the medium. Stated differently, the medium, in the absence of signal energy, is prebiased to one extreme -- the "off" state. Note also that the medium is virtually always contaminated with environmental noise sources; the sun, for example, produces a tremendous amount of in-band IR radiation, resulting in elevated receiver random noise levels. Fluorescent lamps, particularly those incorporating the new electronic ballasts, generate modulated in-band IR, contributing strong spectral components into the hundreds of kilohertz, unnervingly near the lower edge of the signaling [modulation] spectrum.

Three Candidates

While one could certainly assemble a far more inclusive list of encoding schemes than is offered herein, it was deemed appropriate, for several reasons, to limit the discussion to Manchester, PPM, and RZ. Manchester encoding is a widely employed technique, familiar to many for its incorporation in the IEEE 802.3 standard, and PPM has been specifically recommended as a code for optical data links [Ref. 1], particularly where battery operation demands low average transmitter input power. This discussion is intended to be not an exhaustive investigation of all possibilities, but rather a vehicle for the presentation of a specific alternative response to the encoding question, and a context within which that response may be framed. In keeping with this scope and purpose, arguments will appeal more to the reader's intuitive understanding than to his or her technical expertise.

The actual mechanism of coding is quite straightforward in all cases. Manchester encoding, one of several variants of biphase encoding, produces a wavetrain in which each source bit is represented by a transition; the direction of this transition defines the bit state, as shown in Fig. 1a. [A variation encodes source bits as the presence or absence of output transitions.]

The term "PPM" is nonspecific; for the sake of this discussion, the model proposed by Richard Allen of Wireless Research [Ref. 1] will be used. Each four-bit time interval is divided into sixteen equal time slots; a pulse in any one time slot uniquely defines one four-bit source nibble; see Fig. 1b.

The third alternative, RZ, encodes source bits as the presence or absence of a pulse at the beginning of each output bit interval. Fig. 1c depicts these pulses as having a width of one fourth of the bit interval, but this is not mandatory, as will be later shown.

Manchester

There is much to like about Manchester encoding. Certainly the technique seems attractive in several respects: it presents a constant envelope; it has a nonvarying DC component; it offers an abundance of edges for clock extraction; and any number of LSI controller devices, intended for Ethernet applications, will readily support it. One wishes to suggest, though, that the technique, in the context of IR communication, is not without flaw.

Clock extraction is not as trivial as the surfeit of signal transitions might suggest. A cursory examination of a sample Manchester encoded bit stream makes clear the fact that the edges which convey timing information are not consistently of like polarity, implying that the fundamental spectral component of the encoded signal experiences frequent phase reversals. The signal, as such, resists recovery of the clock; indeed, a simple PLL will not reliably phase lock to such a wavetrain. Preliminary processing is required and, especially in cost-sensitive designs, this will commonly take the form of either differentiation or squaring.

Differentiation, whether analog or digital, produces a narrow pulse coincident with each edge of the encoded signal, from which clock edges of consistent polarity may be selected; sadly, this technique is preferential to noise and, whilst perhaps adequate in comparatively noiseless environments [as over a shielded cable], is inapplicable to conditions involving marginal signal-to-noise ratios. Squaring the encoded bitstream is a better means of securing timing transitions, but is more costly than differentiation; certainly, far more sophisticated [and expensive] methods of clock recovery are possible as well.

Having met the challenge of clock extraction, one must employ the resultant information for the reclamation of the original data stream, and here becomes evident a greater disadvantage of the Manchester encoding technique. All bits of source data are encoded with an identical amount of energy; that is to say, within every encoded bit interval, exactly half the time is spent in the "high" state, and half in the "low" state. Because each encoded bit is allocated equal energy, discrimination of source bits must be based on timing alone [unless each half-bit is treated as a separate symbol]; the problem with this is that timing, in practice, is rather easily corrupted, particularly when bit transfer rates ascend to the 4 Mbps region. System bias distortions, multipath effects, and external and internal random noise sources can induce substantial edge jitter, but a Manchester encoded signal can tolerate edge displacements of no more than 1/4 bit time from the nominal; standard Ethernet controller devices will typically tolerate even less, being made deliberately sensitive to aberrations that might indicate the occurrence of collisions.

The advantage of the nonvarying DC content of a Manchester data stream is undone somewhat by the unipolar signaling medium, for a receiver whose frequency response does not extend to DC, or nearly to DC, will express an exponential baseline drift [-- the DC content, though constant over the length of an encoded frame, is compelled by the unipolar medium to be *nonzero*]. Clearly, such behavior need not be ruinous, but in some situations, particularly those in which a short preamble is desirable, it may prove annoying. If the receiver is designed to include baseline restoration of the signal, a reasonably simple addition, the potential for trouble is negated.

Pulse Position Modulation

The case for PPM was put forth in the paper cited in Ref. 1. The technique is undoubtedly superior in terms of average transmitter input power requirements, taking advantage of the error response commonly implemented in these systems: corrupted frames are simply discarded, and degree or extent of corruption is of no consequence. Signaling theory suggests that, all else being equal, the lesser symbol energy per bit associated with PPM must be reflected in a higher bit error rate; however, since within a given frame, two bit errors, or eight, or one hundred, are not worse than a single bit error, the lower energy per bit does not matter; the *symbol* error probability, not the bit error probability, is the parameter of greatest importance.

The claim for the signal-to-noise advantage of PPM is, unfortunately, founded on the mistaken idea that the receiver photodetector will operate into a matched impedance, thereby maximizing power transfer; as discussed earlier, this does not occur in practice. Even so, the higher peak pulse amplitudes permitted by the necessarily small pulse duty cycle do reduce the error probability of a very simple threshold type data regenerator, but convey little benefit to more sophisticated recovery schemes. Whatever value derives from the high pulse amplitudes is not, in any case, unique to PPM.

Several other attributes are unique to this encoding method, though not particularly to its credit. As with Manchester encoding, source bits of both senses are represented with precisely equal energy, and only time serves to differentiate between them, but the required timing resolution is much finer, and very demanding, especially in the presence of noise; whilst a Manchester signal could tolerate as much as 1/4 bit time edge position error, PPM will accommodate no more than half that amount. Both clock and data recovery from a PPM encoded signal can be trying, and powerful techniques for managing poor signal-to-noise ratios, integration foremost among them, are very difficult to apply.

It has been claimed that PPM, like Manchester, has a nonvarying DC content; this is true, but only when considering a sample interval of many symbol times. Clearly, in the case of a sequence of like symbols, the pulses will assume a periodic nature, and the pulse train will demonstrate a constant DC term. Real data, though, will lead to a loss of periodicity, and a small, variable DC component will be evident if the sample window is too narrow. A much worse situation exists when the pairing of pulses, a data content dependent phenomenon, yields a waveform with a spectral component at 1/8 the bit rate; such a link operating at 4 Mbps will need a receiver passband extending down to 500 kHz, sorely limiting the system designer's ability to filter out electronic lamp ballast noise and the like.

Return-to-Zero

The data encoding technique now to be considered has, like Manchester and PPM, both specific strengths and unique flaws, each to be explored in some detail. RZ encoding is probably, for many observers, not an obvious choice for the IR medium, but there is much to recommend it in this application.

Quite unlike the coding methods already described, source bit states are represented in an RZ bit stream by the presence or absence of a single symbol; the timing sensitivities inherent in Manchester and PPM do not obtain here. While it may be defined either way, it is proposed that the presence of a symbol denote a source "0" bit, and the absence thereof [arguably a symbol in its own right] denote a source "1" bit. The "0" symbol, in turn, is defined as a pulse, not necessarily rectangular, initiated at the beginning of the respective bit interval; its width, it will be explained, may be any reasonable fraction of a bit time less than or equal to 1/2. In the absence of overriding design considerations, a rectangular pulse of 1/4 bit time duration is recommended. Clock recovery is markedly simplified because, in contrast to Manchester encoding, RZ guarantees a phase coherent fundamental spectral component, and *every* rising edge is a clock edge.

A compelling advantage of RZ encoding is its scalability over cost, complexity, performance, and speed. The simplest, least costly approach to data regeneration and clock extraction is surely a simple comparator, employed as a fixed threshold detector; though limited in performance, particularly when noise is significant, for short range links it can be quite satisfactory. A threshold detector will function best when the encoded signal pulses have a high aspect ratio [height/width], for the ratio of *peak* signal to *peak* noise will define performance limits. When the application dictates higher performance, expressed as greater communication range, lower average transmitter power, fewer IREDs, &c., RZ encoding is very amenable to the incorporation of receiver noise integration. In such a circumstance, it may well be desirable to exploit the greater efficiency of IREDs at lower power levels; a relatively low amplitude pulse of 1/2 bit time duration may then be appropriate, and will not sacrifice link sensitivity, for an integrating receiver will respond to pulse *energy*, the power-time product, rather than to pulse amplitude alone. It should be noted that, despite the differences between the low and high performance systems just posited, the two are quite capable of intercommunication, the net performance being defined by the characteristics of the minimal design, as would be anticipated.

The market will undoubtedly demand ever faster data transfer rates, and in this, as well, RZ holds an advantage. The unmatched latitude permitted in selection of pulse width and pulse shape, and the 1/2 bit time tolerance of timing variations, will readily facilitate the migration to higher bit rates. As promising as RZ seems to be, however, it is by no means unflawed. Two problems, one minor in the context of the other coding techniques described, and one rather more substantial, require explanation.

The lesser fault derives from the fact that an RZ encoded signal includes, as does a PPM signal, a variable DC component; as mentioned earlier, the incorporation into the receiver of a baseline restoration circuit makes this generally innocuous. Not quite so venial is the fact that an RZ bit stream representing a long sequence of "1" bits is actually no bit stream at all, for it contains no energy. This implies that the spectrum of the encoded signal extends all the way to DC, and that a receiver is at risk of losing bit synchronization with the transmitter; neither condition is tolerable. There exists, fortunately, a simple means of masking this defect: zero bit insertion, perhaps more commonly known as Bitstuffing.

Bitstuffing is a technique, utilized in all HDLC/SDLC controllers, whereby excessively long runs of consecutive "1" bits are fragmented through the addition of non-message "0" bits. In HDLC/SDLC applications, five "1" bits are considered excessive, but for IR communication links, it is proposed that three be taken as the criterion, assuring that the lower edge of the coded signal spectrum will not reach much below 1/4 the bit rate -- not quite as good as Manchester, but a definite improvement on PPM. Such a proposal raises an obvious concern for signaling overhead, but it can be shown that the average bitstuffing load, for uniformly distributed random data, would be 6.25%. The worst case load [for the message body only -- the preamble and the frame delimiting flags are never subject to bitstuffing] is 25%, a most unattractive number, but a most unlikely one as well. Increasing the limit for consecutive "1" bits to four or five in an attempt to reduce the bitstuffing overhead will reduce the average loading to 3.13% or 1.56%, respectively, and the worst case loading to 20% or 16.67%, respectively, but will require greater receiver bandwidth for processing.

The determination of encoding method for infrared signaling will impact strongly not only the performance, and hence marketability, of optical networking technology, but its range of application as well. While Manchester encoding and Pulse Position Modulation may offer certain technical advantages in other contexts, it is submitted that Return-to-Zero encoding can, of the three, best satisfy the diverse needs of the market, flexibly and efficiently supporting a broad continuum of performance and cost levels. RZ can effectively mask over many of the limitations of the transmitting and receiving devices presently available, and will take good advantage of the improved capabilities of devices still to come. The benefits, one submits, make a compelling argument in its favor.

References:

1. Richard Allen, "Infrared Wireless Networks," *IEEE P802.11/91-33* and Richard Allen, "Draft Strawman Infrared PHY Interface Specification," *IEEE P802.11/91-51*.