Determining Optimum Modulation for Inter-Satellite Communications Systems

Gary Mitchell
AeroAstro, Inc.
20145 Ashbrook Pl.
Ashburn, VA 20147
(703) 723-9800 X121
gary.mitchell@aeroastro.com

Abstract. In the small satellite world, it is becoming increasingly popular to use a constellation of small satellites to do the work previously performed by one large satellite. Implicit in this scenario is the requirement for clusters of small satellites to communicate among themselves via an inter-satellite link (ISL).

Traditionally, an ISL has taken the form of a time-division-multiple-access (TDMA) communications system. In a TDMA system, each satellite gets the use of the channel (i.e. gets to transmit) at a particular time and for a particular length of time. Usually this system uses binary-phase-shift keying (BPSK) or some other type of coherent modulation. The choice of BPSK is made because it is more power-efficient than other types of modulation, such as frequency-shift keying (FSK).

When using a BPSK system, there are inherent delays that occur when one satellite's time slot ends and another's begins. These delays are caused by the time required for the BPSK receivers to coherently lock to the incoming BPSK signal. These delays, which cause a direct reduction in system throughput, have to be factored into the overall system design.

It is well known that FSK provides essentially instantaneous communication (i.e. no lock-up time). The penalty for using FSK, however, is the higher transmitter power required compared to BPSK.

As TDMA systems become more crowded, meaning more users to be provided time slots within a given time period, the lock-up time becomes a significant portion of the total available time. This corresponds to an overall reduction in system data throughput. At some point, the advantages of BPSK are outweighed by the required lock-up time when the user changes.

This paper quantitatively examines the tradeoffs involved in selecting a modulation type when the above parameters are considered. A set of metrics is developed to assist in determining the optimum modulation type and multi-variable charts are presented that show the optimum modulation to use for a given scenario.

Introduction

Communications between the satellites in a multiple-satellite cluster requires that either frequency or time be divided between the users. The first option, dividing available frequencies, is referred to as Frequency Division Multiple Access (FDMA). The second option, dividing time among the users (assigning time slots) is referred to as Time Division Multiple Access (TDMA).

The selection of FDMA or TDMA is usually predicated on the relative availability of Radio Frequency (RF) bandwidth. If unlimited RF bandwidth is available, then FDMA is the obvious choice, because of its simplicity. All that is required is that each satellite be assigned a separate channel on which to transmit. The number of channels required generally increases as n, where n is the number of satellites.

In the real world, RF bandwidth is always a scarce resource. Therefore, it is common to use TDMA as the method for providing access to the communications channel for each satellite. This paper addresses some considerations in choosing the optimum type of modulation for TDMA systems.

There are essentially two types of digital communications systems, coherent and non-coherent. Coherent systems, such as Binary Phase Shift Keying (BPSK) systems have an advantage in ultimate system sensitivity. That is, they require the least amount of power to support a given data throughput and error rate.

Non-coherent systems, such as Frequency-Shift Keying (FSK) are poorer performers in terms of the required received power to support the same data throughput and error rate as a coherent system[1]. This often leads to the conclusion that a coherent system is more desirable, because power is always a serious system design constraint in satellite communications systems.

A coherent system has one disadvantage compared to a non-coherent system; it requires a local reference frequency to be generated that is coherent in phase and frequency with the received signal. This involves a phase-locked loop of some type, and that implies a loss of communications time while the loop acquires and locks to the incoming signal.

This “dead time” in coherent systems, which non-coherent systems do not experience, represents a net loss in data throughput because the link is not available 100% of the time. As the number of users in a coherent TDMA system increases this loss in data throughput becomes an increasingly large fraction of the total available data throughput.

A way of comparing the two types of modulation is to consider the radiated RF power to be fixed. In this analysis, if there is no dead time, the coherent system has an advantage in data rate over the non-coherent system. As the dead-time increases, this advantage decreases (because the net data throughput decreases) and at some point the non-coherent modulation becomes more advantageous. This comparison will be used below to develop criteria for choosing one system over the other.

In the sections that follow the dead time for BPSK systems will be calculated for different combinations of communications link data rate and the required system update rate. This information will be presented in a series of figures as the amount of data the system can process as a percentage of the total amount of data the system could process if there were no dead time. When the throughput of the coherent system degrades to 50% or less, non-coherent modulation is more efficient. Also shown will be the availability of the coherent data link, so that, even when coherent communications is more efficient, the amount of data that can be processed can be determined.

TDMA Operation

TDMA operation implies that each satellite has an assigned time slot during which it transmits. Typically, every satellite has a chance to transmit its information for a fixed amount of time and then use of the channel passes to another satellite. The amount of time for all the satellites to have a chance to transmit is referred to herein as the system cycle time (Figure 1). Cycle time can be interpreted as the system update rate, and this time will vary according to the system requirements.

This method implies that each satellite has a common time reference, possibly a GPS clock. Time can also be passed, to a limited degree of precision, over the TDMA network. This, however, requires an initial "bootstrap" protocol so the system can self-determine its configuration and timing. There are other TDMA methods of varying complexity, but this is the only one analyzed here.

Ideally, each satellite would be able to transmit useful information during its entire time slot. In a non-coherent system, this is essentially the case. In a coherent system, there is a certain amount of dead time during each time slot (Figure 2). The total amount of lost system throughput time is the sum of all these dead times. This total will vary with the data rate of the system and the required cycle time.

Coherent and Non-Coherent Modulation

In systems using coherent modulation, such as BPSK, the data is contained in the instantaneous phase of the received signal. To extract this information, a local replica of the transmitter's reference phase must be generated. There are a number of methods for doing this; what they all have in common is that there is a time delay inherent in

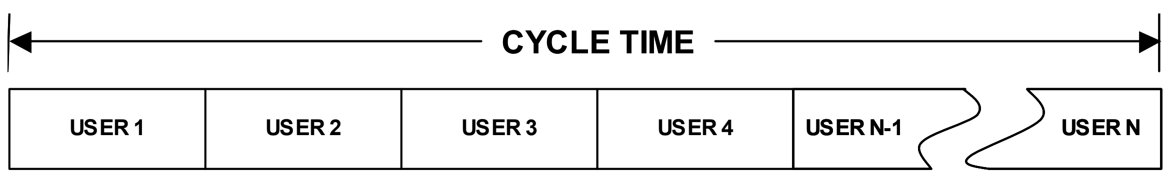

Figure 1 Ideal TDMA System Timing

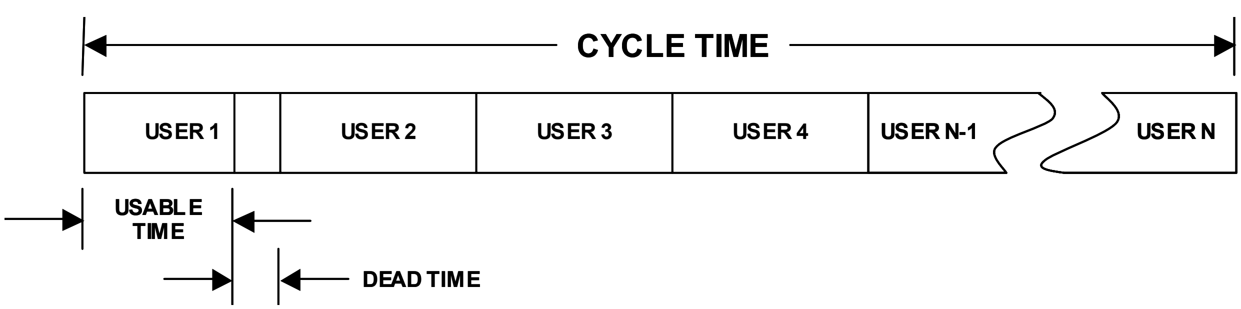

Figure 2 Non-Ideal TDMA System Timing

regenerating the local carrier whenever the system starts. Since the receiver essentially "re-starts" with every TDMA transmission, this can add up to a significant amount of total time.

In non-coherent systems, the information is contained in a signal parameter that can be measured instantaneously, such as amplitude or frequency. Most non-coherent digital system use FSK, so the information is embedded in the instantaneous frequency of the received signal. This can be measured immediately at system start-up, with no delay.

Non-coherent systems have the disadvantage in that they require 3 dB more power to operate at the same level of performance as coherent systems[1]. In some TDMA scenarios, however, this 3 dB disadvantage is more than compensated for by the instantaneous operation.

BPSK Demodulation

BPSK demodulation, as mentioned above, requires a locally generated carrier signal that is both frequency and phase coherent with the frequency and phase of the received signal, if that received signal were not modulated. The problem is that the BPSK modulation process (as generally defined) leaves no spectral component at the modulated signal center frequency. Therefore, one must be generated locally.

There are two commonly used methods for generating this reference, the Costas loop and the carrier-recovery loop. They have similar performance; the carrier-recovery loop is examined here.

Carrier Recovery Loop

In the carrier recovery loop, (figure 3) the received signal, at a convenient downconverted intermediate frequency (IF), is passed through a squaring process. This is typically a multiplier or mixer. This mathematical squaring process generates, along with some other terms, a spectral line at exactly twice the IF.

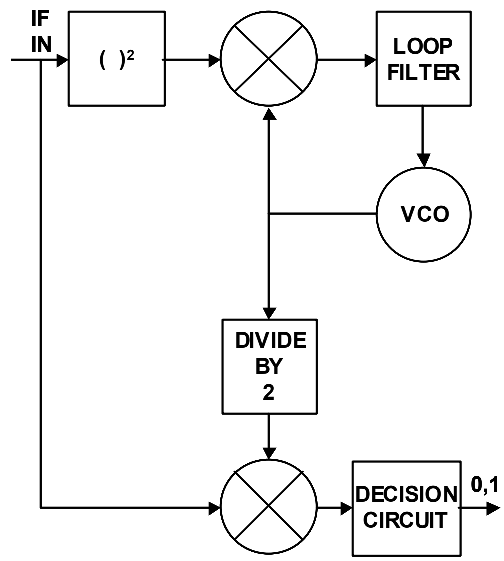

Figure 3 Carrier Recovery Loop

This signal, at twice the IF, is compared in the phase detector (shown in the diagram as a mixer or multiplier) to the phase of a Voltage Controlled Oscillator (VCO) operating at or near this doubled frequency. The phase output of the phase detector is used to steer the VCO until it is exactly phase and frequency locked with the doubled IF signal. Dividing the VCO frequency by two then generates the reference signal required for demodulation.

Multiplication of the original IF signal by this locally generated reference signal generates, after low-pass filtering, a voltage indicating either a zero or one. The decision circuit (typically a voltage comparator) makes the 1 or 0 determination.

The section of the demodulator containing the VCO, phase detector and loop filter is a phase-locked loop. It is the closedloop bandwidth of this loop that determines the response time of the circuit.

FSK Demodulation

As mentioned above, an FSK system carries its information in the instantaneous frequency of the received signal. In a binary (BFSK) system, one frequency corresponds to a 1 and the other to a 0. Demodulation requires only the detection of one frequency or the other.

FSK Demodulator

An FSK demodulator can take several forms, probably the most basic of which is shown below (Figure 4). The IF signal is split into two components, each of which is filtered by a bandpass filter. Each of the two bandpass filters is centered at the frequency that corresponds to either 1 or 0.

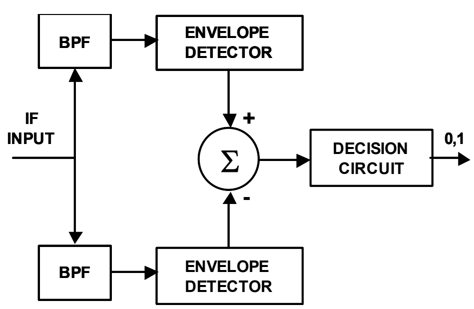

Figure 4 FSK Demodulator

The output of the bandpass filters is then detected using a simple envelope detector. Since there will be power at one frequency and only noise at the other, there will be (relatively) more power at the output of one envelope detector than the other. The two outputs are compared and a 1 or 0 decision made. Since there is no phase-locked loop with its associated delay involved, operation of this circuit is essentially instantaneous.

Coherent Modulation Dead Time

The time delay inherent in coherent demodulation is essentially the time required for the demodulator phase-locked loop to achieve lock. During the time the loop is not locked, there is no local carrier replica to demodulate the incoming signal, so communications cannot occur during this time.

The coherent TDMA system must allow for this dead time each time the transmitting satellite changes. Assume that satellite n is transmitting. The receiving satellites are all locked onto this signal and demodulating data. When the end of the time slot occurs, satellite n stops transmitting and satellite $n+1$ starts transmitting. Since all the system's receivers are currently locked to satellite n they must now acquire satellite $n+1$, which will, in general, differ in both phase and frequency.

Since any individual satellite is now receiving a signal with any frequency offset up to the limit imposed by the quality of all the satellite reference oscillators and with random phase, the demodulator must acquire both frequency and phase. The system must allow for the worst case, which is all receivers having maximum offset and random phase. The dead time is the time required to acquire the frequency and lock the phase.

Phase-Locked Loop Acquisition Time

The phase-locked loop acquisition time has two components; the time required to acquire the frequency and the time required, once the frequency has been acquired, for the loop to achieve phase lock. These two times are referred to as T_{acq} for frequency acquisition time and T_s for phase-lock (or settling time). For a noiseless input signal, these two terms are given by[2]:

$$T_{acq} \cong 3.5 \bullet \frac{(\Delta f)^2}{B_n^3} \qquad (1)$$

$$T_s \cong \frac{1.5}{B_n} \qquad (2)$$

Where:

T_{acq} = Time for the demodulating loop to acquire frequency

T_s = Time for demodulating loop to settle in phase

Δf = Frequency error of received signal

B_n = Noise bandwidth of demodulating loop

The loop acquisition and settling time is seen to be completely dependent upon two terms, frequency offset (Δf) and noise bandwidth (B_n).

Frequency Offset

The frequency offset that the demodulator must acquire is the accuracy of the satellite reference oscillator multiplied to the transmitter output frequency. Since a commonly used inter-satellite band is the 2200-2300 MHz band, a transmitted frequency of 2250 MHz was assumed in the analyses that follow. For example, if the satellite reference oscillator has a specified accuracy of 1 ppm (part-per-million), then the transmitted frequency can be in error by as much as 2250 Hz.

Noise Bandwidth

The noise bandwidth, B_n, of the phase-locked loop is a function of the system data rates and signal to noise ratios (SNRs), and cannot be set arbitrarily. If the bandwidth is too narrow, the loop will take longer than necessary to lock. If set too wide, there will not be sufficient SNR to drive the loop into lock.

The acquisition and settling time shown above is for a noiseless input signal. This is clearly not the case in any useful communications system. An interesting result is shown below (Figure 5)[2]. In interpreting this graph, it seems that SNRs above 20 dB do not contribute to the ability of the loop to lock. It also seems that SNRs below 20 dB are not sufficient for the loop to lock in a reasonable amount of time. Therefore, 20 dB is assumed herein to be as good as the noiseless case in terms of loop acquisition and lock times.

A common requirement for communications links is that they transmit data with a bit error rate (BER) of one bit in 10,000. This BER, for a BPSK system, requires a signal bandwidth SNR of 9.6 dB[2] (10 dB is used herein for simplicity). This means that in the signal bandwidth, the SNR must be at least 10 dB. Assuming this requirement is satisfied by design of the system, the noise bandwidth, B_n, of the demodulating loop can be calculated. For a required loop noise bandwidth SNR of 20 dB and a signal bandwidth SNR of 10 dB provided, the loop noise bandwidth is the signal bandwidth divided by 10.

For example, if a communications system is to support a data rate of 50 kilo-bits per second (kbps) it must provide an SNR of 10 dB in a bandwidth of 50 kHz. This then sets the optimum demodulating loop bandwidth, B_n, at 5 kHz. In each of the cases analyzed below, this is the SNR used.

Coherent Modulation Link Availability

Analyses were performed for a number of cases to determine the link availability of a

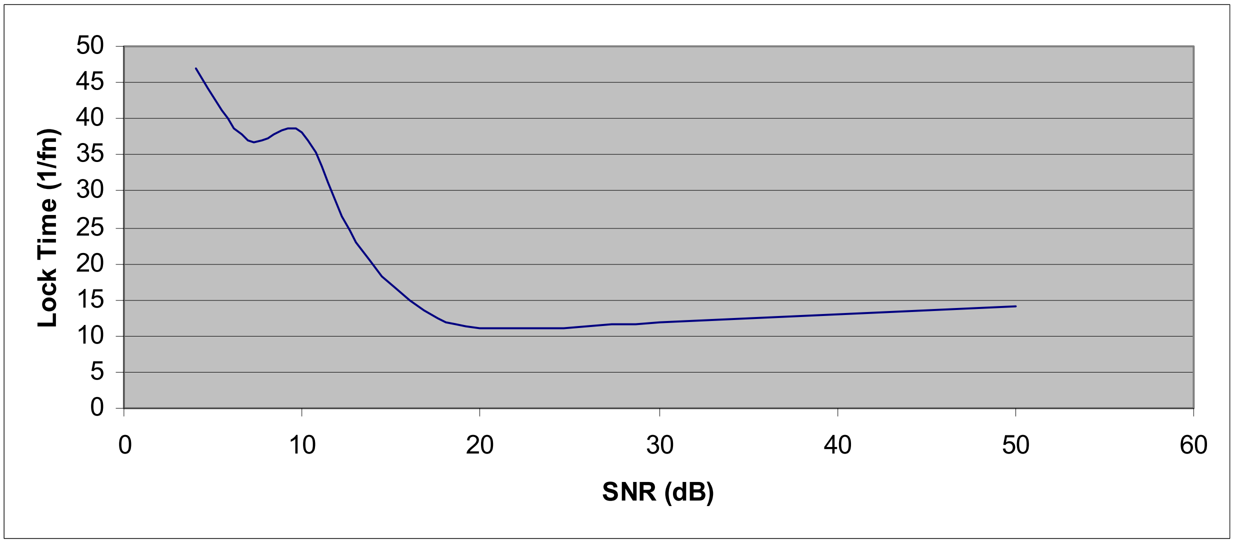

Figure 5 Phase-Locked Loop Lock Time

BPSK communications system. In each of the cases analyzed, the frequency offsets and loop bandwidths were calculated as developed above. The data rates, number of users and cycle times of the system were all allowed to vary. Cycle time can be considered to be the inverse of the system update rate.

In each case, the total dead time ($T_{acq} + T_s$) for one user time was divided by the allocated user time. The availability of the link is defined as:

$$Availability = (1 - DeadTime)$$

The results of these analyses are shown below for a number of interesting cases (Figures 6-9). For a particular cycle time, link availability is plotted versus system data rate. Each of the curves within a graph represents a particular number of users between 4 and 128. Figure 10 shows link performance for a fixed data rate with varying numbers of users by system cycle time.

In each of the cases below a frequency error of 1 ppm was assumed for the reference oscillator error. It should be noted that this is the combined error of both the sending satellite and the receiving satellite. For these results to be valid, the system must be specified so that the error of all oscillators is 0.5 ppm or better.

Symbol Synchronization

In addition to demodulation of the received data, the receiver must also provide a clock that is synchronized with the incoming data. This generally (but not always) requires another phase-locked loop. The requirement to provide a clock signal was not included in these analyses because both types of modulation require this function and the focus of this paper is the difference in performance between the two types of modulation.

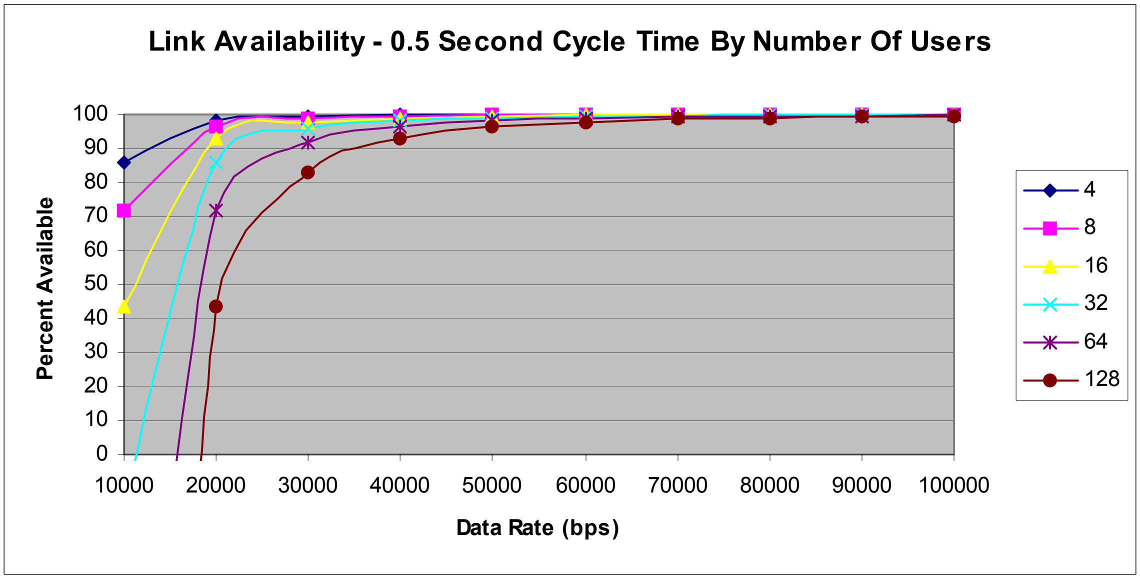

Figure 6 Link Availability – 0.5 Second Cycle Time

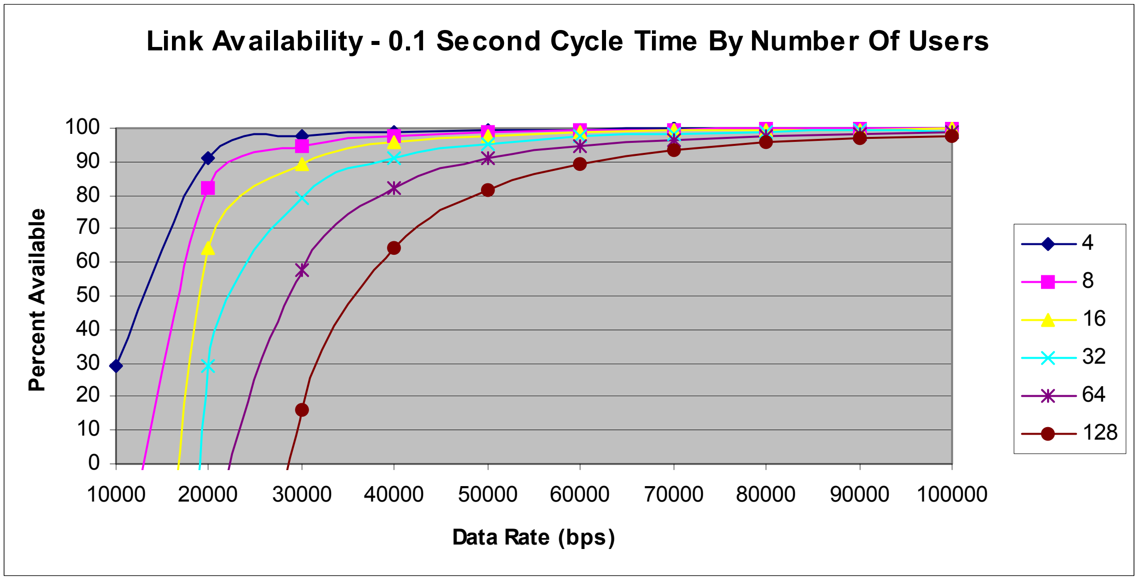

Figure 7 Link Availability – 0.1 Second Cycle Time

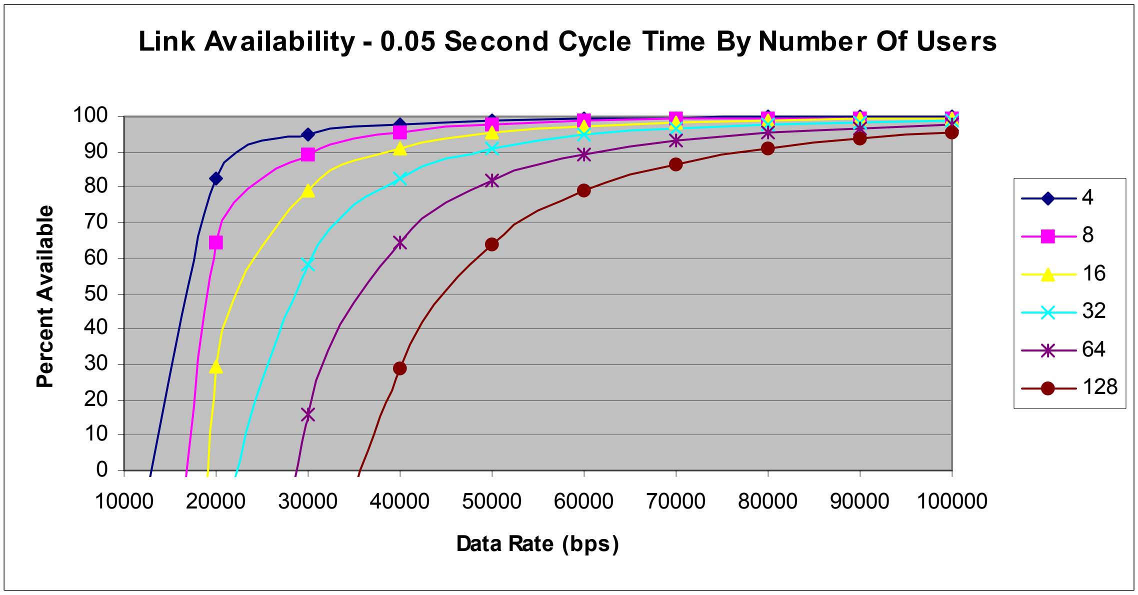

Figure 8 Link Availability – 0.05 Second Cycle Time

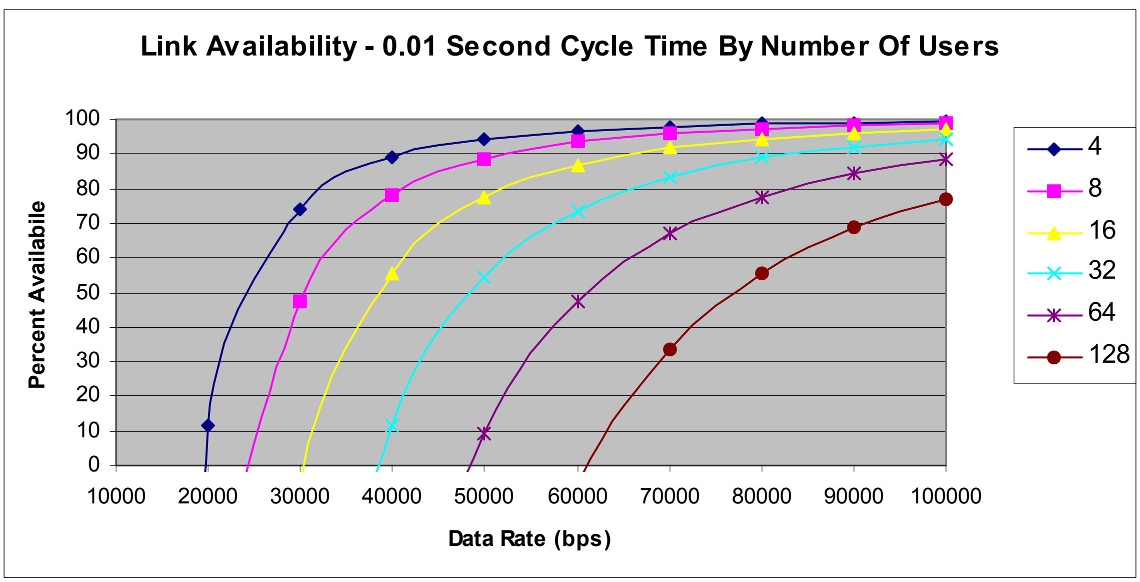

Figure 9 Link Availability – 0.01 Second Cycle Time

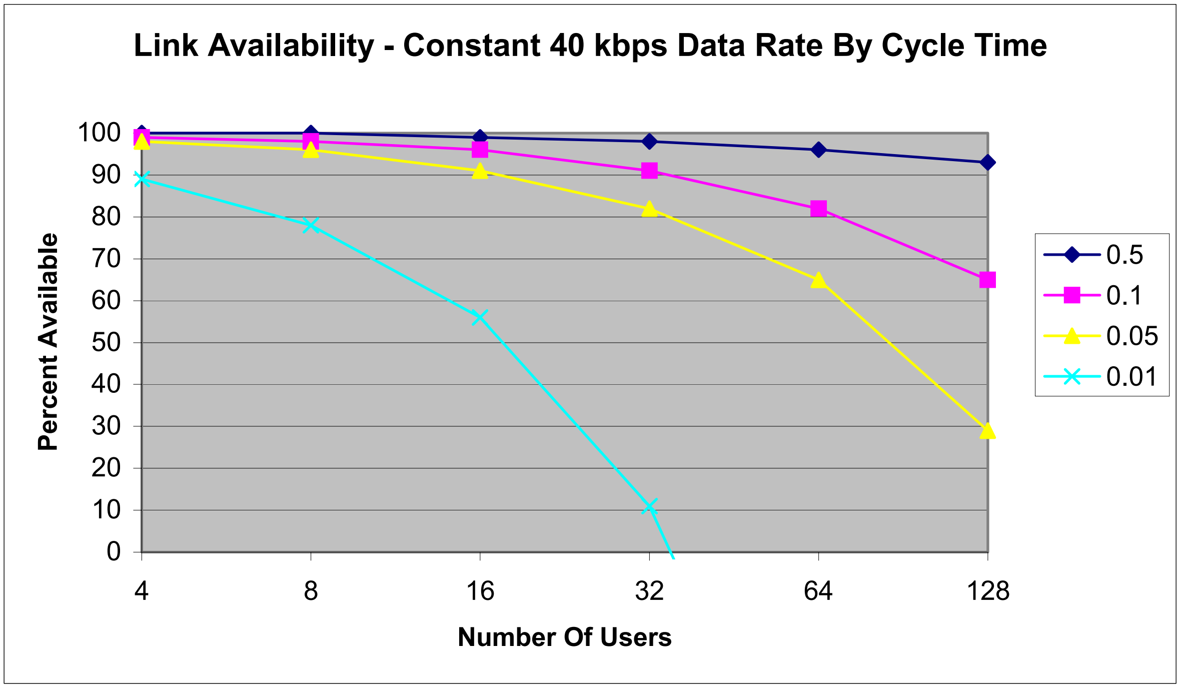

Figure 10 Link Availability At Constant Data Rate

Conclusions

This paper has shown that the availability of a BPSK communications to function can vary greatly because of the time required for the demodulating loop to phase lock to the received signal. In each case shown above, when the availability of the link drops below 50%, the use of a non-coherent system should be considered, based on the idea that a BPSK with 100% availability has twice the performance of a BFSK system.

Also shown is that, for a BPSK system, the communications link availability may become so low as to be useless for communications. The set of conditions analyzed represent a limited number of possible communications systems. The information contained in this paper should allow the analysis of other cases of interest.

References:

1. Smith, David R., Digital Transmission Systems, Van Nostrand Reinhold, New York, 1993.

2. Spilker, James J., Digital Communications by Satellite, Prentice Hall, Englewood Cliffs, NJ, 1977.