IAA-01-IAA.9.1.09

Error Correction Scheme in Active Seti

Y. Dutil
ABB Bomem Inc.

S. Dumas

IAA-01-IAA.9.1.09

ERROR CORRECTION SCHEME IN ACTIVE SETI

Yvan Dutil
ABB Bomem Inc.
585 Charest blvd East, Suite 300
Québec, Qc., Canada
G1K 9H4
Yvan.Dutil@sympatico.ca

Stéphane Dumas
79 Paul-Borduas
Loretteville, Qc, Canada
G2A 3M5
Stephane_Dumas@sympatico.ca

Abstract

By opposition to passive SETI, which want to detect extra-terrestrial civilizations, the goal of active SETI is to establish a communication link with those civilizations. Already difficult, this task is complicated by the lack of knowledge about the nearest civilization. In addition, technical and astrophysical considerations severely limit the communication bandwidth for radio waves. The construction of a message, which can be decoded by an extra-terrestrial civilization as been described in details by the Dutch mathematician Hans Freudenthal. Based on these general principles, Frank Drake created a signal, which was sent from Arecibo. However, it appears now that additional steps are needed in a careful design to facilitate its decoding.

Introduction

Historically, in the study of problems related to communication with extra-terrestrial life little attention has been paid to the noise in the communication channel.

In general, the analysis is limited to the calculation of the signal to noise ratio at the receiver:

$$\frac{S}{N} = \frac{PA_R}{4pR^2WkT} \qquad (1)$$

where P is the radiated power by the source, A_R is the collecting area of the receiving antenna, R is the distance from the source, W the bandwidth of the signal and T the noise temperature of the system.

This equation has been applied to the calculation of the maximum communication range for a transmitter. Most of the time, a signal to noise ratio of one is assumed. This very low value leads to optimistic range estimation. However, a low signal to noise ratio has other more rational benefices.

Indeed, a communication system has a maximum error-free channel capacity given by:

$$C = W \log_2 \left(1 + \frac{S}{N} \right) \qquad (2)$$

Where W is the bandwidth, S is the signal power and N the noise power[1].

Obviously, wider the bandwidth more efficient is the communication. This leaded some authors to consider that advanced civilizations may use wide band signal to communicate with us[2]. Unfortunately, such wideband signal would be nearly impossible to differentiate from noise. Therefore, narrow band signal is expected at least to establish the first contact[3,4].

The other way to increase the communication channel efficiency is to reduce the signal to noise ratio. However, lowering the signal to noise ratio also increases the bit error rate. At very low signal to noise ratio, the message can be totally undecipherable. Beyond the simple repetition of the message *ad infinitum*[5], there is other ways of fighting the noise. In particular, error-correcting scheme could be imbedded in the message.

Interstellar scintillation handling

Over galactic distances, the interstellar medium induces some scintillation and modulation of electromagnetic signals. This modulation is higher at lower frequencies and is a function of the sky direction. Over large distances, the depth of the modulation can exceed 100%, making the signal very difficult to decode. Even if the modulation often amplifies the signal, in average, it attenuates it and reduces the probability of detection. The only regime where the gain could offset the loss is when the signal is just below the detection threshold[6,7].

The effects of the interstellar scintillation can be alleviate in part by a clever modulation strategy. Taking advantage of the variation of the interstellar scintillation with wavelength, the signal can be sent on a frequency comb in such way that each wavelength experiences a different modulation[8].

However, using more lines dilutes the power density and, in consequence, the possibility of detection. As pointed by Cordes & Sullivan, the optimal number of lines increases with the signal to detection threshold ratio, which is very closely related to the signal to noise ratio[6].

However, using two lines cost nothing when a frequency shift keying (also known has FM) is used. Since each of the two lines is the complement of the other, producing two observations of the same signal for the price of one[8].

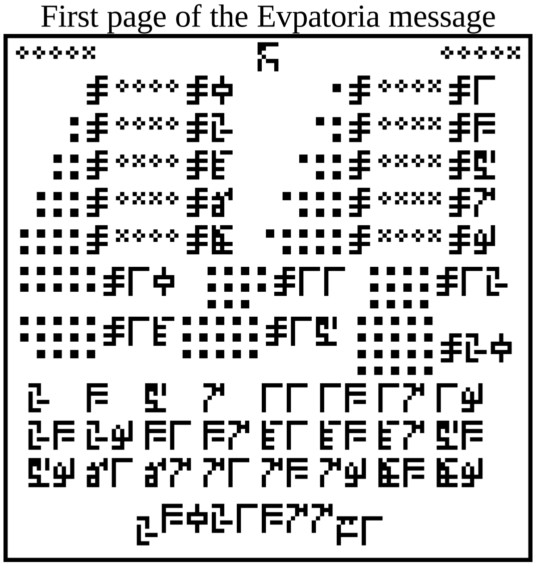
Cordes & Sullivan neglected this fact and under-estimated the performance of the two lines modulation scheme. In consequence, if we correct their result, it can be shown that a two-lines signal recovers most of the detection lost due to the scintillation for all signal to detection threshold ratio and still benefits from the amplification effect for sub-threshold events. Accordingly, utilization of two lines and frequency shift keying appears to be the best approach to deal with the scintillation effect.

Nevertheless, there are additional benefits of choosing the multi-lines transmission. First, it increases the signal resistance to radio-frequency interference and, if the receiver does not cover all the radio bands instantaneously, it increases the probability of detection.

<u>Noise resistant format</u>

The first and critical step to increase the noise resistance is to insure that the format will be easily detected and interpreted by the reader. One format approach proposed for interstellar messages was to use the product of two prime numbers to construct an image. Unfortunately, this method works only if all the bits are present. As an example, the message send by Frank Drake from Arecibo in 1974 did not have any feature to support some noise degradation[9]. Even before its transmission, it was already reported that if one bit was lost, the message could become unreadable[10].

Figure 1
First page of the Evpatoria message



However, there is a simple trick to protect the format of fax-like transmission: synchronization marks. This may be as simple as taking a pause between each line has proposed by H. W. Nieman & C. Nieman in 1920[11]. For the Encounter 2001 project, we have chosen to begin and end each line with a "1" (represented by a black pixel, see figure 1)[12]. In fact, the actual synchronization sequence is "0110". Since each line is 127 bits long (we used a prime number more for its esthetic value than its usefulness), we have sacrificed 3.15% of our bandwidth only to carry the format. Even in presence of noise, this synchronization signal can be easily detected by the presence of a "piquet fence" pattern in the Fourier spectrum (Dirac comb) or by the classical cryptographic Kappa test[13,14].

In order to improve the logical structure of the message, we have divided it in 23 pages of 127 lines (again prime number were used for esthetic reason). Each page has been delimited by a frame (see figure 1). In addition, at the top of each page, a header is placed with the page number on each sides of an ideogram describing the subject of the page.

However, this format is somewhat restrictive. Indeed, to maximize the usage of the bandwidth each page should be as filled as possible. Unfortunately, this is often impossible if one wants to keep only one subject per page. In some circumstance, we have been forced to do not respect the logical order of idea to use efficiently the available space. For example, we have put the list of the prime numbers on the page 1 even if the exponential notation needed to write the largest prime known is only explained later. With the experience, it appears now the page length should be variable in order to maximize the overall throughput of the transmission while keeping the width constant.

In term of communication efficiency, the Evpatoria transmission was far from being optimal. Indeed, it was essentially a monochromatic signal spiced with a supplement of information. Over the 370 967 bits sent, 314 239 was “1” and 56 768 was “0” (5.54 times more “0” than “1”). Since we used a frequency shift keying modulation scheme most of the time the signal will be on the “0” frequency. In addition, “0” tend to be send in long stretch (white line). All of this leads to a very small modulation.

In the frequency space this is translated as a concentration of the power near the carrier. Indeed, for the band “0”, 57% of the total power is concentrated within a bandwidth of 0.01 Hz and 65.2% within 1 Hz. However, for the band “1”, the value are 5.6% within 0.01 Hz and 26.3% within 1 Hz. The difference between the two bands comes from average shorter length of the black lines.

This concentration of energy in narrow band eases the detection of the signal at larger distance. Nevertheless, the fractional with and black dots should be nearer to 1:1, in order, to maximize the utilization of the available bandwidth. How it could be done simultaneously with the conservation of the narrow bandwidth characteristics needs to be investigated further.

Orthogonal characters set

For a purely logical message as proposed by Hans Freudenthal each element of the logical proposition can be represented by its own “tone”[16].

For such purely logical message, basic telecommunication theory shows that those signals should be made of equal blocks of bits, and each unique block is represented by a different orthogonal (uncorrelated) signal. This encoding method maximizes the efficiency of the communication link by allowing the maximum rate of transmission with the minimum error rate[17].

However, in practice such pure logic message appears to be ruled out by most experts. In place, a hybrid communication scheme including logical proposition and graphical representation is preferred[18-25]. In the hybrid representation, the idea is to have a character per proposition. In that context, we shall speak of ideogram, not characters.

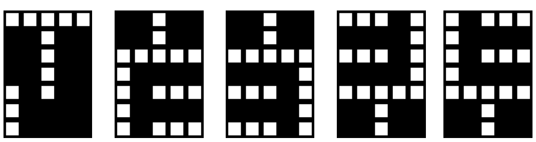
It should be noted that choosing ideograms instead of binary sequence already offers some improvement against noise resistance. In fact in a fax-like transmission, ideogram will be spread on many lines. This increases its resistance against short burst of noise like radio frequency interference or interstellar scintillation.

The choice of the graphical representation of the ideograms could be arbitrary, since graphical representation is strongly culturally biased (synthetic characters offer the additional advantage of being culturally neutral). However, in presence of noise, the situation is slightly different. Common sense and basic communication theory tells you that ideograms must be as different as possible from each other.

It would be quite easy to generate directly an arbitrary list of ideograms from an orthogonal binary representation. Appropriate binary sequence could be generated by any standard error correcting scheme[26]. However, specific constrains of the hybrid message coding forced us to find another solution.

For example, it is necessary to assure the orthogonality between ideograms independently of the orientation (see figure 2). This requirement comes from the fact that the orientation changes the meaning of some characters in natural language (ex.: “()”, “<>” or “dbpq”). In order to avoid any misunderstanding characters have to be orthogonal and this whatever their orientation.

Figure 2
Orthogonal character set properties



Another, but weaker, requirement is to keep a uniform shape of the ideogram. In many languages, small symbols have a diacritic function and therefore a change in size of the ideogram may create some confusion. To keep the size constant each ideogram has to keep at least one bit in the “1” state along each extreme column and line.

To generate those characters a brute force method was applied. We simply scanned all the possible characters and kept only those fulfilling the previous requirement for geometry. From this subsample, we only kept those with a minimal difference with any other characters previously selected. This technique is not optimal however, and we have discovered later that it was much more efficient to randomly select the character instead of scanning them sequentially.

This minimal difference between each character is also known as Hamming distance. And from the communication theory, it is well known that a Hamming distance of 2n+1 is necessary in order to correct an error of n bits[27].

For the 1999 edition, the Hamming distance has been fixed at 7 bits. Accordingly, an error of 3 bits could be corrected. This is equivalent to a bit error rate of 8.6%, a monstrous error rate by telecommunication standard! The binary notation used for page numbering is even more robust, since its Hamming distance is equal to 9 over 9 bits. In theory, it should be able to handle a bit error rate of 44%! At that point however, it is not obvious that an intelligent observer would attempt any decoding!

A caveat must be established, in SETI applications, things are not that simple since the receiver does not own the list of correct codes like in the usual communication applications. In consequence, the receiver must build the list by itself, even before starting the actual decoding. To do so, it will have to use some kind of cluster analysis algorithm[28]. In such circumstance, some false characters might be created and some good characters might be undetected due to the action of noise fluctuations. However, part of these errors might be recoverable from the analysis of the context.

Some might fear that the inclusion of the noise resistance might come at the cost of the reduction of the useful bandwidth. However, this is not an issue for two reasons. First, as we have shown previously, fax like transmissions tend to underuse the available bandwidth. Second, using orthogonal coding scheme is more efficient than repeating the message.

For example, to produce a distance of 7 bits by a simple bit repetition, the binary sequence would need to by recopied 7 times. Therefore, to encode 128 characters (a number similar to the one used in the Evpatoria message in 1999) with this Hamming distance someone would need 49 bits (7x7) if it simply recopies the bits. By contrast, the same result can be achieved with an orthogonal coding scheme for a larger number of ideograms with only 35 bits (28.6% improvement) even if the additional requirements we have added might have increase the number of bits needed. Further optimizations of the algorithm can even improve this figure.

Further optimizations

Another argument against the used of Reed-Solomon like algorithm is the non-uniform probability distribution of word in natural language. Indeed, in natural language, letters, words, phoneme and morphemes follow a power law distribution[29]. This is also true for dolphin whistles and for DNA sequences[30,31]. The universality of this power law distribution appears to be a direct consequence of the optimization of language for efficient communication[32].

A statistical analysis of the message sent from Evpatoria in 1999 shows that it is also the case for an anti-cryptographic message [the data are best fit by the modified Zipf law n=493/(r+1)^0.99]. This is somewhat surprising due to the highly mathematical nature of the message.

From the knowledge of this statistical behavior, we could, like in natural language, adopt longer bit chain to code rare words. Alternatively, we may choose characters such as to maximize the average distance between them.

We attempted to do such optimization. However, our result appears modest since we have only managed to increase the average Hamming distance from 13.65 to 13.8 bits. This poor result is somewhat deceiving since the character set has not been design with such optimization scheme in mind. Therefore, we are still optimist about the possibility of creating an optimal generating function that will efficiently use the statistical properties of the message.

Further improvements can be imagined. For example, the character set may also be optimized in order to handle some temporal resolution degradation of the receiver end.

Conclusion

We have demonstrate that noise resistance should be taken into account in active SETI and than this can done easily with scarifying any bandwidth. Since the requirements for such error correction scheme comes directly from the basic communication theory it is likely to be known by any advance civilization willing to communicate. By symmetry, researchers involved in passive SETI should evaluate its impacts on their detection scheme.

Further researches are needed to match the coding scheme to statistical properties of the Zipf law. One possibility would be to used smaller character for numbers than for others ideograms. In addition, an end to end simulation, including the decoding of the message should be attempted in order to test our coding scheme.

We would like to thanks Jill Tarter, who has encouraged us to publish our work and Douglas Vakoch, who provided us with rare scientific papers to support our research. We would like to thanks also Jean-Claude Leclerc and Jean-Francois Rivard for their useful comments and suggestions.

N.B.: This work is a personnal initiative of Y. Dutil without any support from ABB Bomem Inc.

1. Shannon, C. E., 1948, “A mathematical theory of communication”, Bell System Tech. J., 27, pp 623-656

2. Shostak, S., 1995, “SETI at Wider Bandwidths?”, in Progress in the search for Extraterrestrial life, ASP conf, 74, p 447., G. Seth Shostak ed.

3. Drake, F., 1960, Sky & Telescope, 19, pp 140-142

4. Drake, F. D., & Helou, G., 1978, “The Optimum Frequencies for Interstellar Communications as Influenced by Minimum Bandwidth”, NAIC 76, Cornell University, New York

5. Petrovich, N. T., 1997, “A SETI Correspondent Helps us to Discover their Signals, Lost in the Noise of our Receivers”, Ap&SS, 252, 59-66

6. Cordes, L. M., Sullivan, W. T., 1995, “Astrophysical Coding: A New Approach to SETI Signals”, in Progress in the Search for Extra-terrestrial life, ASP conf. Series, vol 74, p 325, G. Seth Shostak ed.

7. Cordes, J. M., Lazio, T. J. W., & Sagan, C., 1997, “Scintillation-induced Intermittency in SETI”, ApJ, 487, 782

8. Cohen, N, Charlton, D., 1995, “Polychromatic SETI”, in Progress in the Search for Extra-terrestrial life, ASP conf. Series, vol 74, p 313, G. Seth Shostak ed.

9. The Staff at the National Astronomy and Ionosphere Center, 1975, “The Arecibo Message of November, 1974”, Icarus 26, 462-466

10. Shklovskii, J. S., & Sagan, C., 1966, “Intelligent life in the Universe”, Holden-Day, San Francisco

11. Nieman, H. W. & Nieman, C. W, 1930, “What Shall We Say to Mars?”, Scientific American, 122, 298, 312

12. Details technical description of the message can be found on this web page (in French and English): http://pages.infinit.net/lachapel/seti/introduction.html

13. Deavours, C. A., 1977, “The Kappa test”, Cryptologia, 1, 3, 60-67

14. Deavours, C. A., “Extraterrestrial communication: A cryptologic perspective”, in Extraterrestrials: Science and Alien intelligence, ed. Edward Regis Jr., Cambridge University Press, pp 201-204

16. Freudenthal, H., “Lincos: Design of a language for a cosmic intercourse, part 1” (Amsterdam, North-Holland Publishing, Compagny, 1960).

17. Jones, H. W., 1995, “Optimum Signal Modulation for Interstellar communication”, in Progress in the Search for Extra-terrestrial life, ASP conf. Series, vol 74, p 369, G. Seth Shostak ed.

18. Arbib, M. A., 1979, “The psychology of Interstellar Communication”, Cosmic Search, vol 1, summer, pp 21-24

19. Vakoch, D., 1978, “Possible pictorial messages for communication with extra-terrestrial intelligence”, Journal of the Minnesota Academy of Science, 44, 23-25

20. Vakoch, D., 1998, “Constructing Message to Extraterrestrial: An Exosemiotic Perspective”, Acta Astronautica, 42, 697-704

21. Vakoch D, 1998, “The Dialogic Model: Representing Human Diversity in Messages to Extraterrestrials”, Acta Astronautica, Vol. 42, N°10-12, pp.705-710

22. Vakoch D., “Signs of Life beyond Earth: A Semiotic Analysis of Interstellar Messages”, Leonardo, Vol. 31, N°4, pp.313-319, 1998

23. DeVito, C., Oehrle, R. “A Language Based on the Fundamental Facts of Science”, Journal of the British Interplanetary Society, 1990, Vol. 43, pp. 561-568

24. De Vito, C., 1992, “Language Based on Science”, Acta Astronautica, 26, 268

25. De Vito, C, 1992, “Language, Science and the Search for Extraterrestrial Intelligence”, Leonardo, vol. 25, 1, p 13-16

26. Reed, I.S. & Solomon, G. 1960 “Polynomial codes over certain finite fields”, SIAM Journal on Applied Mathematics, 8, 300-304.

27. Hamming, R.W., 1950, “Error Detecting and Error Correcting Codes”, Bell System Technical Journal

28. Tryon, R.C., 1939, “Cluster Analysis: Correlation Profile and Orthometric (Factor Analysis for the Isolation of Unities in Mind and Personality)”, Ann Arbor, Michigan, Edward.

29. Zipf, G. K., 1949, “Human Behavior and the Principle of Least effort”, Addison-Wesley press

30. McCowan, B., Hansert, S F. & Doyle, L. R. 1999, “Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires”, Animal Behaviour, 57, 409-419

31. Mantegna, R. N. *et al* 1994, “Linguistic Features of Noncoding DNA”, Physical Review Letter, 73, 3169-3172

32. Mendelbrot, B. , 1951, “Adaptation d’un message à la ligne de transmission.”, Comptes Rendus de l’Académie des Sciences, (Paris), 232, 1638-1640 & 2003-2005