Upper bound on false alarm rate for landmine detection and classification using syntactic pattern recognition

Ahmed O. Nasif, Brian L. Mark, Kenneth J. Hintz, and Nathalia Peixoto

Dept. of Electrical and Computer Engineering
George Mason University
Fairfax, VA 22030, U.S.A.

ABSTRACT

Recently, there has been considerable interest in the development of robust, cost-effective and high performance non-metallic landmine detection systems using ground penetrating radar (GPR). Many of the available solutions try to discriminate landmines from clutter by extracting some form of statistical or geometrical information from the raw GPR data, and oftentimes, it is difficult to assess the performance of such systems without performing extensive field experiments. In our approach, a landmine is characterized by a binary-valued string corresponding to its impedance discontinuity profile in the depth direction. This profile can be detected very quickly utilizing syntactic pattern recognition. Such an approach is expected to be very robust in terms of probability of detection (P_d) and low false alarm rates (FAR), since it exploits the inner structure of a landmine. In this paper, we develop a method to calculate an upper bound on the FAR, which is the probability of false alarm per unit area. First, we parameterize the number of possible mine patterns in terms of the number of impedance discontinuities, dither and noise. Then, a combinatorial enumeration technique is used to quantify the number of admissible strings. The upper bound on FAR is given as the ratio of an upper bound on the number of possible mine pattern strings to the number of admissible strings per unit area. The numerical results show that the upper bound is smaller than the FAR reported in the literature for a wide range of parameter choices.

Keywords: Landmine detection, Ground penetrating radar, False alarm rate

1. INTRODUCTION

There is a continued interest in the development of a robust, cost-effective and high performance landmine detection and classification system. A variety of approaches have been discussed in the research literature.[1,2] Most approaches make a distinction between two major classes of landmines based on their metal content. There are those which contain a sufficient amount of metal to be detected by electromagnetic induction (EMI) detectors and those which are essentially non-metallic. For non-metallic landmines, the most common sensor used for detection is ground penetrating radar (GPR)[3,4] and the most common detection methodology is *anomaly detection*, sometimes colloquially called *"blob"* detection. This methodology is typically implemented utilizing a pre-screener followed by a more detailed analysis of the GPR data for detection leading to a two-pass operational approach. Improvements to the anomaly detector have recently been made which fuse EMI and GPR data as well as take into account contextual information by alternative methods of analyzing the GPR data before selecting the algorithm appropriate to the soil type and environment. Anomaly detection typically depends on computing running statistics which characterize the clutter background and select for further analysis those regions whose statistics differ from the background statistics by some threshold value.

An alternative approach to anomaly detection and its variants is the application of syntactic pattern recognition (SPR)[5] . In SPR, single columns of range data from the GPR (A-Scan) are processed so as to convert the signal into a binary-valued string of the same length as the original data. This binary-valued string is comprised of zeros except for where there are changes in impedance. The binary-valued string is then presented to a finite state machine (FSM) language recognizer for detection. This is a single-pass operation which does not require a pre-screener. It is also extremely fast in execution since the FSM operates

at one clock cycle per state transition, namely one clock cycle per range bin, independent of the number of states in the FSM.

The binarization of the A-scan approach was chosen because it is well known that the radar cross section (RCS) measured by a radar looking at a complex target is highly variable with aspect angle. We have the distinct advantage of being able to assume that most landmines are oriented approximately parallel to the earth yielding a repeatable and recognizable pattern. While this slight variation in either the orientation of the landmine or the relative geometry between the GPR and the landmine contributes to amplitude variations in the measured RCS with range, it has been demonstrated that the locations of impedance discontinuities can be reliably detected through preprocessing followed by a binarization procedure. That is, the binary-valued string representing the location of impedance discontinuities of a prototypical buried landmine vary only a slight amount which is well within the capabilities of being recognized as a language by an appropriately designed FSM.

The key metrics of interest for a landmine detection system are: (1) probability of detection (P_d), and (2) false alarm rate (FAR). The FAR is defined as the probability of false alarm per unit area, which for landmine detection, is a more meaningful performance measure than simply the probability of false alarm. Given that we have a suitable method for the preprocessing and production of the binary-valued A-scan, it remains to recognize the string associated with a landmine and discriminate it from other strings produced by clutter and noise. We call the process of producing the binary-valued A-scans a *binarization* scheme. It is a fundamental requirement of language recognizers that they not only recognize all the words in the language but no others.

In the computation of the false alarm rate (FAR) of an SPR language recognizer, one needs to know two numbers. The first is the number of strings that represent the language of the landmine. While it may seem that one can have a language consisting of a single string, operationally this is not possible since the statistical variations in the A-scan of a landmine present themselves as small movements of impedance discontinuities about their nominal locations. These variations are small and enumerable and it is not difficult to compute the number of strings representing variations on a landmine's characteristic string and this computation is discussed in Section 3. The second number required for FAR estimation is the total number of all *admissible* strings. Admissible strings are those which can be produced by the binarization process. It may first appear that for a landmine string of length n, that there would be 2^n possible strings produced by the binarization process, some of which would be as a result of the landmine and some as a result of clutter. A simple example shows that this is not true. If we assume that the binarization process cannot produce adjacent impedance discontinuities, then strings of the form "$\cdots 0110 \cdots$" are not possible. This limitation, among others, means that the number of admissible strings is less than the 2^n possible strings in a string of length n. It is for this reason that a methodology based on combinatorial enumeration has been developed in this paper, namely to be able to compute the number of admissible strings as a first attempt to put a bound on the FAR of SPR as well as provide guidance in the development of the binarization process. We note that in the SPR approach, such a bound can serve as a rough estimate of the expected performance in terms of FAR, whereas in most traditional approaches one needs to perform detailed field tests to evaluate the system's FAR.

The remainder of the paper is organized as follows. In Section 2, we characterize the binary sequence space due to an arbitrary binarization scheme using a combinatorial enumeration technique. In particular, we compute the total number of admissible strings for a fixed string length. In Section 3, we compute an upper bound on the FAR. Section 4 presents some numerical results showing the expected FAR performance of the SPR approach. We then use the results from Section 2 and contrast them, in Section 5, against statistics of binary strings obtained from real landmine data. In Section 6 we discuss the significance of the experimental and theoretical results to our future work in detecting landmines with the syntactic pattern recognition method.

2. CHARACTERIZATION OF THE BINARY SEQUENCE SPACE

Suppose we are given a random sequence, $\{X_i \in \mathcal{R} : 1 \leq i \leq n\}$, where n denotes the length of the sequence. The sequence $\{X_i\}$ could represent, for example, a sequence of measurements comprising an A-scan obtained

using GPR to detect the presence of landmines. It could also represent a sequence which is generated as a result of performing some processing on the raw data of a GPR A-scan. To perform further processing of the measurements using the SPR approach, the random sequence $\{X_i\}$ is converted into a binary sequence $\{B_i \in \{0, 1\}\}$.

There are various ways to convert a given sequence* $\{x_i\}$ to $\{b_i\}$. As mentioned before, we refer to such a conversion process as *binarization*. For example, we can have the following simple binarization rule:

$$b_0 \triangleq 0, \tag{1}$$

$$b_i = \begin{cases} 1, & \text{if } x_i > x_{i-1} \text{ and } x_i > x_{i+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

for $i \geq 1$. In words, b_i is set to one if x_i is a *local maximum* with respect to x_{i-1} and x_{i+1}. Otherwise, b_i is set to 0. Another scheme may be to consider all *local extrema* of $\{x_i\}$. Obviously, we can have more sophisticated binarization schemes. In this section, we study the characteristics of the deterministic properties of the binary sequence space that results from application of an arbitrary binarization rule, subject to the constraint that no consecutive "1"s are allowed.

The sequence $\{b_i\}$ may be viewed as a binary string with certain properties. Let $\mathcal{B}$ denote the set of binary strings that could possibly correspond to a sequence $\{b_i\}$ obtained by applying a binarization rule to a finite or infinite sequence of values $\{x_i\}$. In other words,

$$\{b_i\} \in \mathcal{B},$$

for all binary sequences $\{b_i\}$ generated by the binarization method. Furthermore, any sequence $b \in \mathcal{B}$ corresponds to a valid binary sequence $\{b_i\}$ obtained from the binarization process.

A given sequence $v \in \mathcal{B}$ may be characterized as a binary string that begins and ends with "0" and does not contain the string "11". We refer to such a sequence as an *admissible* string. We can compactly represent the set $\mathcal{B}$ by the regular expression (see Appendix A)

$$\mathcal{B} = (00^*1)^*00^*. \tag{3}$$

Let $\Phi(x)$ denote the generating function of the set $\mathcal{B}$ with respect to sequence length (see Appendix B). In particular, define the weight function

$$w(\sigma) \triangleq \text{length of } \sigma, \quad \sigma \in \mathcal{B}.$$

Then

$$\Phi(x) = \sum_{\sigma \in \mathcal{B}} x^{w(\sigma)}.$$

From (3), we can obtain the following expression for $\Phi(x)$:

$$\Phi(x) = \left[1 - \frac{x}{1-x} \cdot x\right]^{-1} \cdot \frac{x}{1-x} = \frac{x}{1-x-x^2}. \tag{4}$$

Let $\mathcal{B}_n$ denote the set of strings of length n that belong to $\mathcal{B}$:

$$\mathcal{B}_n \triangleq \{b \in \mathcal{B} : |b| = n\}. \tag{5}$$

The following proposition gives a closed-form expression for the number of admissible strings of length n in $\mathcal{B}$. A proof is given in Appendix C.

---

*We will follow the convention that lowercase letters represent deterministic variables, whereas uppercase letters represent random variables.

PROPOSITION 2.1 (ADMISSIBLE STRINGS). *The number of admissible strings of length n in $\mathcal{B}$ is given by*

$$|\mathcal{B}_n| = \sum_{j=\lfloor \frac{n}{2} \rfloor}^{n-1} \binom{j}{n-j-1}. \tag{6}$$

Hence, the fraction of binary strings of length n that belong to the set $\mathcal{B}$ is given by

$$f_n = \frac{|\mathcal{B}_n|}{2^n} = 2^{-n} \sum_{j=\lfloor \frac{n}{2} \rfloor}^{n-1} \binom{j}{n-j-1}. \tag{7}$$

Fig. 1 shows a plot of f_n vs. n for $n = 1$ to 20. One sees that f_n decreases rapidly as a function of n. In particular, $f_{20} \approx 6.45 \times 10^{-3}$, which means that out of the 2^{20} binary strings of length 20, less than 0.65% of them are in the space $\mathcal{B}$. Note that $f_2 = \frac{1}{2^2}$ and $f_3 = \frac{2}{2^3}$.

It may also be of interest to consider the number of 1's that occur in a string belonging to the set $\mathcal{B}$. Let $w_1(\sigma)$ denote the number of 1's in string $\sigma \in \mathcal{B}$. To address this issue, we consider a bivariate generating function $\Phi(x, y)$, where the coefficient of term $x^n y^m$ represents the number of strings of length n, containing m 1's. In particular, define the weight function

$$w_1(\sigma) \triangleq \text{number of 1's in } \sigma, \quad \sigma \in \mathcal{B}.$$

and

$$\Phi(x, y) \triangleq \sum_{\sigma \in \mathcal{B}} x^{w(\sigma)} y^{w_1(\sigma)}. \tag{8}$$

From (3), we can then obtain the following expression for $\Phi(x, y)$:

$$\Phi(x, y) = \left[1 - \frac{xy}{1-x} \cdot x \right]^{-1} \cdot \frac{x}{1-x} = \frac{x}{1 - x(1+xy)}. \tag{9}$$

Let $\mathcal{B}_{n,m}$ denote the set of strings in $\mathcal{B}$ of length n and containing m 1's. The number of strings in the set $\mathcal{B}_{n,m}$ is given by following proposition. A proof is given in Appendix D.

PROPOSITION 2.2 (ADMISSIBLE STRINGS WITH m 1'S). *The number of admissible strings of length n containing m 1's in $\mathcal{B}$ is given by*

$$|\mathcal{B}_{n,m}| = \binom{n-m-1}{m}, \quad 0 \le m \le \left\lceil \frac{n}{2} \right\rceil - 1. \tag{10}$$

Let $f_{n,m}$ denote the fraction of strings of length n in the set $\mathcal{B}$, containing m 1's. From (10) and (6), we obtain

$$f_{n,m} = \frac{|\mathcal{B}_{n,m}|}{|\mathcal{B}_n|} = \frac{\binom{n-m-1}{m}}{\sum_{j=\lfloor \frac{n}{2} \rfloor}^{n-1} \binom{j}{n-j-1}}, \quad 0 \le m \le \left\lceil \frac{n}{2} \right\rceil - 1. \tag{11}$$

Fig. 2 shows a bar graph of $f_{n,m}$ vs. m when $n = 20$. Note that the distribution has a mode at $m = 5$ and has a relatively symmetric shape about this mode.

3. UPPER BOUND ON FALSE ALARM RATE

As mentioned in Section 1, due to statistical variations it may not be possible to have an unique binary string corresponding to a particular landmine. However, we can introduce some parameters that capture statistical uncertainties. We assume that all valid mine patterns have a known prefix "100". We denote the locations of the "1"s in a valid mine pattern of length n by a vector $\mathbf{q}$. For example, if $\mathbf{q} = (\mathbf{q}(1),\ \mathbf{q}(2),\ \cdots,\ \mathbf{q}(k),\ \cdots,\ \mathbf{q}(K)) = (1,\ 10,\ \cdots,\ 79,\ \cdots,\ L)$, where $K < L \leq n$, it means that the location of the kth "1" of the mine pattern is 79. In the SPR approach, it is possible to represent a particular landmine by $\mathbf{q}$ and its variation according to some parameterization.

We propose to characterize the binary-valued string corresponding to a landmine pattern (of length n) using the following three parameters: (i) the total number of "1"s in the binary string sequence (K, the length of $\mathbf{q}$), (ii) dither or by how much the elements of $\mathbf{q}$ may get shifted (d), and (iii) the number of *noise pulses* (J), that are not contained in $\mathbf{q}$ but may appear erroneously. A mine pattern with dither value d is characterized by a set $\mathcal{Q}_d = \{\mathbf{q} \pm 1, \mathbf{q} \pm 2, \cdots, \mathbf{q} \pm d\}$. By a noise pulse we mean the erroneous change of a "0" bit to "1". We assume that a noise pulse does not occur in any of the positions contained in $\mathbf{q}$, $\mathcal{Q}_d$, the top of the mine, and at mine prefix ("100") locations. The number of valid mine patterns of length n using the parameters (K, d, J) is upper bounded by

$$N_{\text{mine}}^{(u)} = (2d+1)^K \sum_{j=0}^{J} \binom{n - (2d+1)K - 3}{j}. \tag{12}$$

Note that the $(2d+1)$ refers to the number of variations in the string pattern caused by each "1" present in the mine string, and the "3" present in the top argument of the choose function denotes the length of the mine prefix. Clearly, it is possible to have some other criteria or parameters to select the valid mine patterns. But as long as we can evaluate, either analytically or empirically, an upper bound on the number of such valid mine patterns, $N_{\text{mine}}^{(u)}$, our approach of computing an upper bound on FAR would be valid.

Assuming that each string is equally likely to occur, an upper bound on the probability of false alarm is given by the ratio of the number of valid mine pattern strings to the number of admissible strings, $P_{\text{FA}}^{(u)} = \frac{N_{\text{mine}}^{(u)}}{|\mathcal{B}_n|}$. Suppose the footprint area of the GPR corresponding to each A-scan is denoted by D in units of m^2. Since the FAR is defined as the probability of false alarm per unit area we have the following upper bound on FAR:

$$\text{FAR} = \frac{P_{\text{FA}}^{(u)}}{D} = \frac{N_{\text{mine}}^{(u)}}{|\mathcal{B}_n|D}. \tag{13}$$

This upper bound on FAR can serve as a theoretical performance measure of the SPR approach for landmine detection. It can also guide us towards an appropriate binarization scheme such that acceptable FAR values can be achieved.

4. NUMERICAL EXAMPLES

In this section, we provide two numerical examples to illustrate the dependence of the mine pattern parameters on the FAR upper bound. We consider the footprint of a typical GPR, which has area $D = 0.0025$ m^2 and a mine prefix of length 3. In Fig. 3, we plot the upper bound on FAR as a function of string length (n), and the number of 1's in the mine pattern (m), for different values of noise $(J = 0, 1, 2)$ and a dither value of $d = 2$. The surface plot shows that for a fixed string length n, as the number of 1's in the mine pattern m increases, the FAR also increases. This is because the numerator of the bound $(N_{\text{mine}}^{(u)})$ increases with increasing m, while the denominator $(|\mathcal{B}_n|)$ remains unchanged (cf. (6),(12), and (13)). On the other hand, for fixed values of m, the FAR decreases as the string length n increases. This is because $|\mathcal{B}_n|$ increases faster than $N_{\text{mine}}^{(u)}$ as a function of n, causing the bound to decrease. As expected, the bound uniformly increases with increasing noise. In Fig. 4, the bound is plotted as a function of m for fixed values dither and noise, $(d, J) = (2, 2)$, with various values of n. We observe that there is roughly a two orders of magnitude decrease in the bound with every increase in the string length of 10 bits.

5. COMPARISON OF EXPERIMENTAL AND THEORETICAL STRING DISTRIBUTIONS

Once the theoretical approach has been established and argued for, one interesting question remains: to what extent do the experimental results match the numerical data? The implications of such a question for the SPR method, or for any method applied to the identification of a pattern, are crucial: matching distributions may guarantee low false alarm rate, and that guarantee is one main goal of binarization algorithms.

We have used multiple methods for binarization of raw and processed data. As representative samples, here we inspect results obtained with two independent binarization schemes, labeled schemes **b1** and **b2**. Every dataset is examined for strings of length $n \in [20, 140]$, and for strings containing m 1's, with $m \in [2, 13]$. For each combination of (n, m), entire sets of data were surveyed, containing approximately 133 million datapoints (i.e., pixels). For **b1**, over 2 million strings are found to comply with the (n, m) ranges. For **b2**, over 20 million strings are retrieved, indicating a first difference between the two schemes. When graphing these results, the dichotomy is more obvious: the peaks in **b2** (see Fig. 6) are absent in **b1** (see Fig. 5). Furthermore, in **b1** the position of the single peak on the length of the string (n), as if reminiscent of a Rayleigh distribution, is directly proportional to the number of 1's in the string (m). While this is intuitively acceptable, the spiky distributions of **b2** are not.

As stated above, the task of comparing theoretical to experimental results would not be complete without a side-by-side plot. Among the many possibilities of realizing such exercise, we have chosen to fix $n = 80$ (strings of length 80 pixels) and vary m (number of 1's), similarly to the histogram presented in Fig. 2. Figs. 7 and 8 show the resulting distributions of the experimental (blue) datasets for binarization schemes **b1** and **b2**, respectively. Both are compared with the theoretical counterpart (green) (cf. (11)). The experimental data clearly has a different mean compared to the theoretical one, and this is not surprising since the theoretical distribution assumes a uniform distribution of all admissible strings. Such plots can be extremely useful in extracting the mine pattern strings and also assessing the performance of different binarization schemes.

6. CONCLUSION

In this paper, we have studied the performance of the SPR approach to landmine detection and classification in terms of the FAR. In particular, an upper bound on the FAR is proposed, where the bound is given as a fraction of the possible mine pattern strings with respect to the total number of admissible strings that can result from an arbitrary binarization scheme. It is demonstrated that the tool of combinatorial enumeration is very useful in studying the string statistics for syntactic landmine detection.

In this initial study we have made the simplifying assumption that all strings are equally likely, which may not be the case in real landmine data. Therefore, it would be of great interest to relax this assumption in the admissible set, and derive a new upper bound on the FAR. It is also of interest to incorporate the mine pattern parameters in our string enumeration technique. These developments can potentially guide us to a method of extracting landmine string patterns that have very attractive detection and false alarm performances.

ACKNOWLEDGMENTS

APPENDIX A. NOTATION FOR REGULAR EXPRESSIONS

If $a = 0101$ and $b = 1110$ are binary strings, then ab is the string formed by concatenating a and b:

$$ab = 01011110.$$

If A and B are sets of binary strings

$$AB \triangleq \{ab : a \in A, b \in B\},$$

is the set of strings formed by concatenating any string in A with any string in B. We also define $A^2 \triangleq AA$, and in general,

$$A^n \triangleq \underbrace{AA \cdots A}_{n \text{ times}}.$$

Let ϵ denote the *empty string*. By convention, we define $A^0 \triangleq \{\epsilon\}$ as the set consisting of just the empty string. Next we define

$$A^* \triangleq \epsilon \cup A \cup A^2 \cup A^3 \cup \cdots = \bigcup_{i \geq 0} A^i.$$

For brevity of notation, we define $0^* \triangleq \{0\}^*$ and $1^* \triangleq \{1\}^*$.

It is important to note that AB is not, in general, equivalent to the Cartesian product $A \times B$. However, if each of the elements of AB is uniquely generated, i.e., if we can determine uniquely which part of the element came from A and which part from B, then AB is in fact equivalent to $A \times B$. In this case, the *Product Lemma* applies to generating functions defined on the set AB (see Appendix B).

APPENDIX B. GENERATING FUNCTIONS

A generating function of a real-valued sequence $\{a_i\}$ is a formal power series[6]

$$\Phi(x) = \sum_{n \geq 0} a_i x^i. \tag{14}$$

We will use $[x^i]\Phi(x)$ to denote the coefficient of x^i in $\Phi(x)$. Hence, if $\Phi(x)$ is given by (14), $a_i = [x^i]\Phi(x)$.

Given a set S, we associated with each element a nonnegative integer $w(\sigma)$ called the *weight* of σ. Let us define the generating function on S with respect to the weight function $w(\cdot)$ as follows:

$$\Phi_S(x) = \sum_{\sigma \in S} x^{w(\sigma)}. \tag{15}$$

If we write $\Phi_S(x)$ in the form (14), i.e.,

$$\Phi_S(x) = \sum_{i \geq 0} a_i x^i,$$

we see that

$$a_i = \text{number of elements of } S \text{ of weight } i.$$

Two useful rules for manipulating generating functions defined on sets according to (15) are the so-called *Sum* and *Product Lemmas*, which are straightforward to prove.

LEMMA B.1 (SUM LEMMA). *Suppose $S = A \cup B$ and $A \cup B = \emptyset$. Then*

$$\Phi_S(x) = \Phi_A(x) + \Phi_B(x).$$

LEMMA B.2 (PRODUCT LEMMA). *Suppose $S = A \times B$ and for each $\sigma = (a, b) \in S$, $w(\sigma) = w(a) + w(b)$. Then*

$$\Phi_S(x) = \Phi_A(x)\Phi_B(x).$$

As an example, let us represent the set of all binary strings using the following regular expression (see Appendix A):

$$S = A^*,$$

where $A = \{0, 1\}$. Let $w(\sigma)$ denote the length of a binary string $\sigma \in A$. The generating function of the set $\{0\}$ with respect to $w(\cdot)$ is simply x. Likewise, the generating function of $\{1\}$ is x. Therefore, the generating function of A with respect to $w(\cdot)$ is given by $2x$, i.e., $\Phi_A(x) = 2x$. By the *Product Lemma*, the generating function of A^i, for $i \geq 0$, is given by $\Phi_{A^i}(x) = (2x)^i$. By the *Sum Lemma*,

$$\Phi_S(x) = \sum_{i \geq 0} \Phi_{A^i}(x) = \sum_{i \geq 0} (2x)^i = \frac{1}{1 - 2x}.$$

Therefore, the number of binary strings of length n is given by

$$[x^n]\Phi_S(x) = 2^n,$$

as we would expect.

Though this example is rather trivial, it illustrates how the sum and product lemmas can be used to derive an expression for the generating function, and finally, how the coefficient $[x^n]\Phi_S(x)$ can be obtained. The results (4) and (6) are obtained using this approach.

APPENDIX C. PROOF OF PROPOSITION ON ADMISSIBLE STRINGS

The number of admissible strings of length n in $\mathcal{B}$ is given by

$$|\mathcal{B}_n| = [x^n]\Phi(x) = [x^n]\frac{x}{1 - x(1 + x)} = [x^n]\sum_{j \geq 0} x[x(1 + x)]^j \tag{16}$$

$$= \sum_{j \geq 0} [x^{n-j-1}](1 + x)^j = \sum_{j \geq 0} \binom{j}{n - j - 1} = \sum_{j = \lfloor \frac{n}{2} \rfloor}^{n-1} \binom{j}{n - j - 1}. \tag{17}$$

APPENDIX D. PROOF OF PROPOSITION ON ADMISSIBLE STRINGS CONTAINING M 1'S

The number of admissible strings of length n containing m 1's in $\mathcal{B}$ is given by

$$|\mathcal{B}_{n,m}| = [x^n y^m]\Phi(x, y) = [x^n y^m]\frac{x}{1 - xy(1 + x)} \tag{18}$$

$$= [y^m]\sum_{j \geq 0} [x^n]x[x(1 + xy)]^j = [y^m]\sum_{j \geq 0} [x^{n-j-1}](1 + xy)^j \tag{19}$$

$$= [y^m]\sum_{j \geq 0} \binom{j}{n - j - 1} y^{n-j-1} = \binom{n - m - 1}{m}, \quad 0 \leq m \leq \left\lceil \frac{n}{2} \right\rceil - 1. \tag{20}$$

This result can also be derived using a more direct argument.

REFERENCES

[1] D. J. Daniels, "A review of landmine detection using GPR," *Proc. European Radar Conf. (EuRAD'08)*, pp. 280–283, Oct. 2008.

[2] J. N. Wilson, P. Gader, L. Wen-Hsiung, H. Frigui, and K. C. Ho, "A Large-Scale Systematic Evaluation of Algorithms Using Ground-Penetrating Radar for Landmine Detection and Discrimination," *IEEE Trans. Geosci. Remote Sens.* **45**, pp. 2560–2572, Aug. 2007.

[3] H. M. Jol, *Ground Penetrating Radar Theory and Applications*, Elsevier, 2009.

[4] D. J. Daniels, *Ground Penetrating Radar*, MPG Books Ltd., reprint of 2004 edition, UK, 2007.

[5] K. J. Hintz, N. Peixoto, and D. Hwang, "Syntactic landmine detection and classification," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIV*, R. S. Harmon, J. T. Broach, J. H. Holloway, Jr., ed., *Proc. SPIE* **7303**, May 2009.

[6] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Dover Publications, reprint of 1983 edition, 2004.
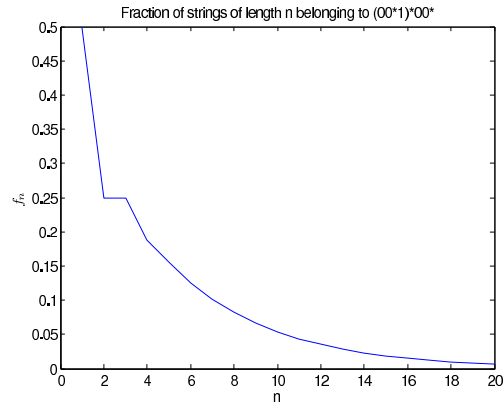
Figure 1: Fraction of strings, f_n, in the binary sequence space $\mathcal{B}$, as a function of string length n.
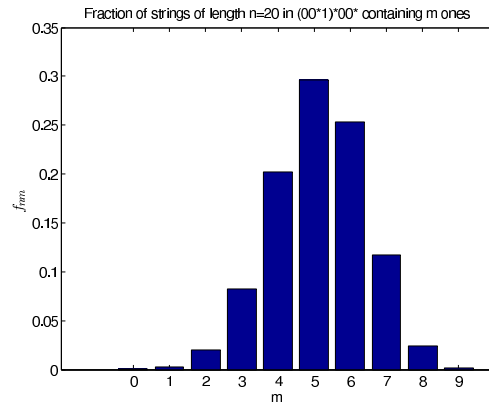
Figure 2: Fraction, f_{nm}, of strings of length $n = 20$ containing m ones in the binary sequence space $\mathcal{B}$, as a function of m.
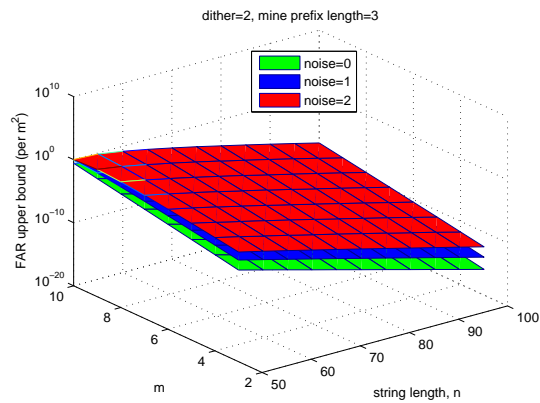

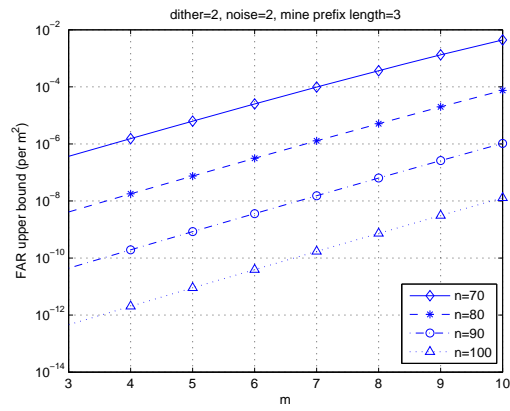Figure 3: FAR upper bound as a function of string length and the number of 1's in the mine pattern, (n, m).

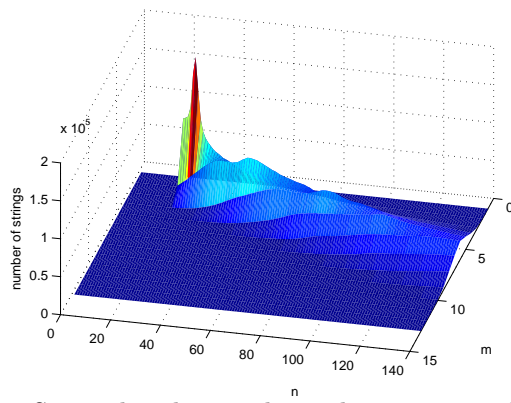Figure 4: FAR upper bound as a function of the number of 1's in the mine pattern, m, for different string lengths.

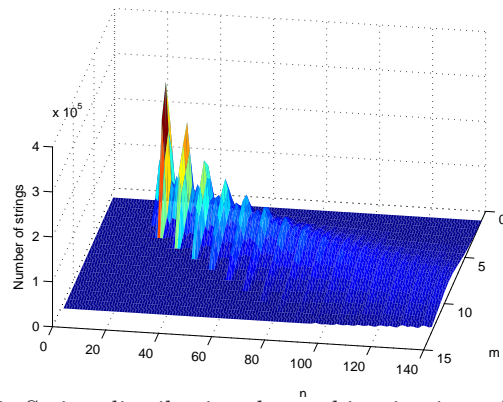Figure 5: String distribution due to binarization scheme **b1**.

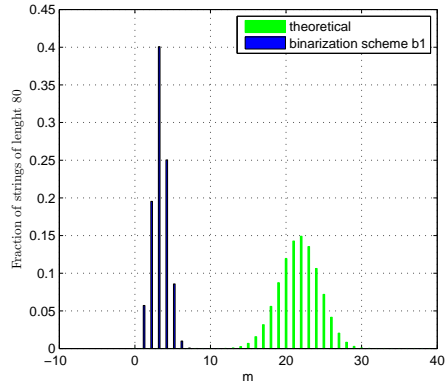Figure 6: String distribution due to binarization scheme **b2**.

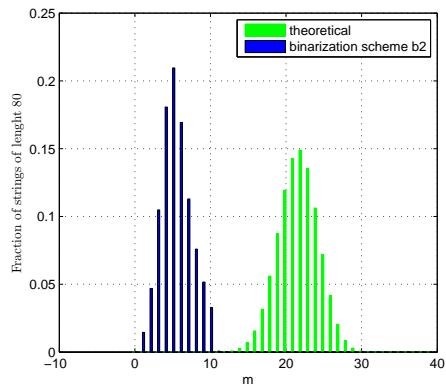Figure 7: Fraction of strings of length 80 using binarization scheme **b1**.

Figure 8: Fraction of strings of length 80 using binarization scheme **b2**.