

Using Block Frequencies to Break Hill Ciphers

Brian D. Douville

briandouville@gmail.com

Department of Mathematical Sciences
Appalachian State University
Boone, North Carolina, USA

Richard E. Klima

klimare@appstate.edu

Department of Mathematical Sciences
Appalachian State University
Boone, North Carolina, USA

Neil P. Sigmon

npsigmon@radford.edu

Department of Mathematics and Statistics
Radford University
Radford, Virginia, USA

***Abstract:** This paper focuses on the breaking of Hill ciphers, a classical encryption method rooted in linear algebra, without knowledge of the key. We will explore the challenges associated with breaking Hill ciphers, and introduce a pragmatic method employing frequency analysis of letter combinations. The initial phase provides a theoretical foundation for breaking ciphers, emphasizing challenges and mathematical complexity specific to Hill ciphers. We introduce a Maplelet that uses frequency analysis of letter combinations, and demonstrate its effectiveness in decrypting messages without a known key. The results highlight the method's effectiveness and unveil vulnerabilities within Hill ciphers. The work also adds a creative perspective to the exploration of cryptanalysis without explicit key knowledge, offering practical insights into breaking classical encryption techniques. This project represents progress in decrypting classical ciphers and creates different methods for imaginative approaches to encryption challenges. The findings provide a basis for further exploration into keyless decryption techniques, fostering creativity in the field of cryptanalysis.*

1 Introduction

In the digital age, with information spread throughout countless networks, the protection of sensitive data has become a problem of paramount importance. Cryptography, the ancient art of concealing messages and modern science of securing communication, stands as the bastion against unauthorized

access. However, as we delve into this paper, our gaze shifts from the intricate mechanisms of cryptographic systems to the captivating world of codebreaking, where the resilience of encryption is put to the test.

The history of cryptography is rife with stories of cryptographic triumphs and defeats, from the decryption of ancient Roman scripts to the breaking of the Enigma code during World War II. The focus of this paper is less on the development and implementation of cryptographic algorithms, and instead more on the theory of cryptanalysis. This paper will also explore techniques that can be utilized to reveal the secrets guarded by cryptographic systems, emphasizing the significance of understanding vulnerabilities for a comprehensive perspective on information security.

2 Hill Ciphers

One way mathematics and cryptology are connected is that in many types of ciphers, the encryption and decryption procedures can or must be expressed using mathematical operations. However, the art of cryptology existed long before its connections with mathematics were observed. William Friedman was one of the first to connect mathematics and cryptology with his index of coincidence, but while Friedman used mathematics in breaking Vigenère ciphers, the encryption and decryption procedures in Vigenère ciphers do not have to be expressed using mathematical operations. The first type of cipher that actually used mathematics in its encryption and decryption procedures appeared in a 1929 article entitled *Cryptography in an algebraic alphabet* [1], written by a little-known mathematician from Hunter College named Lester Hill.

We will consider plaintext messages expressed using only the letters in the alphabet A, B, C, . . . , Z, and convert these letters into numbers using the correspondences A = 0, B = 1, C = 2, . . . , Z = 25. In educational settings, *affine ciphers* are often proposed as a precursor to Hill ciphers. An affine cipher uses a calculation of the form $y = (ax + b) \bmod 26$, or, equivalently, $y = (xa + b) \bmod 26$, for some a and b in $\mathbb{Z}_{26} = \{0, 1, 2, 3, \dots, 25\}$, where a is chosen so that $a^{-1} \bmod 26$ exists. We give a thorough description of affine ciphers in [3].

Consider affine ciphers for which $b = 0$, so that encryption is done with a calculation of the form $y = xa \bmod 26$. Hill ciphers are a generalization of these affine ciphers, where the plaintext x is a row matrix \mathbf{x} containing a list of plaintext numbers, and a is a square matrix A with the same number of rows as the number of entries in \mathbf{x} . That is, for a Hill cipher, we begin by grouping the plaintext numbers in order into row matrices¹ of some fixed size, say $1 \times n$, and then choosing an encryption matrix A of size $n \times n$ with entries in \mathbb{Z}_{26} , where A is chosen so that $A^{-1} \bmod 26$ exists. Then, for each plaintext row matrix \mathbf{x} , we form a corresponding ciphertext row matrix \mathbf{y} with a calculation of the form $\mathbf{y} = \mathbf{x}A \bmod 26$. The entries in the ciphertext row matrices \mathbf{y} , when strung together in order and converted back into letters, form the ciphertext.

More specifically, let A be a 2×2 matrix for which $A^{-1} \bmod 26$ exists, and suppose we wish to encrypt a plaintext using a Hill cipher with the encryption matrix A . We begin by converting the list of letters in the plaintext into a list of numbers, and then grouping these numbers in order into row matrices of size 1×2 . Note that the last plaintext row matrix will be filled only if the length of the plaintext is a multiple of 2 (or, more generally, a multiple of the number of rows in A). If this is not the case, then the plaintext can and should be padded at the end with a letter (or letters, possibly, if A were larger than 2×2) so that the length of the plaintext is a multiple of 2. Assuming this, suppose

¹With Hill ciphers, it is possible to use column matrices here instead of row matrices. However, since using column matrices does not increase (or decrease) the security of Hill ciphers, for consistency we will only use row matrices.

the plaintext is of length n , with plaintext numbers $x_1, x_2, x_3, \dots, x_n$ in order. Then the corresponding list of ciphertext numbers $y_1, y_2, y_3, \dots, y_n$ is found by forming the following matrix products.

$$\begin{aligned} \begin{bmatrix} y_1 & y_2 \end{bmatrix} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} A \pmod{26} \\ \begin{bmatrix} y_3 & y_4 \end{bmatrix} &= \begin{bmatrix} x_3 & x_4 \end{bmatrix} A \pmod{26} \\ \begin{bmatrix} y_5 & y_6 \end{bmatrix} &= \begin{bmatrix} x_5 & x_6 \end{bmatrix} A \pmod{26} \\ &\vdots \\ \begin{bmatrix} y_{n-1} & y_n \end{bmatrix} &= \begin{bmatrix} x_{n-1} & x_n \end{bmatrix} A \pmod{26} \end{aligned}$$

The ciphertext numbers $y_1, y_2, y_3, \dots, y_n$ are then converted back into letters to give the final ciphertext.

For example, consider a Hill cipher with the following encryption matrix A .

$$A = \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix}$$

To use this matrix to encrypt the plaintext BE HERE AT SEVEN, since this plaintext contains 13 letters, we begin by padding an A at the end of the plaintext so that its length will be a multiple of 2. Next, we use the correspondences A = 0, B = 1, C = 2, ..., Z = 25 to convert the plaintext from a list of letters into a list of numbers.

B	E	H	E	R	E	A	T	S	E	V	E	N	A
1	4	7	4	17	4	0	19	18	4	21	4	13	0

To encrypt the plaintext, we form the following matrix products.

$$\begin{aligned} \begin{bmatrix} 1 & 4 \end{bmatrix} A &= \begin{bmatrix} 1 & 4 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 6 & 5 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 7 & 4 \end{bmatrix} A &= \begin{bmatrix} 7 & 4 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 2 & 11 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 17 & 4 \end{bmatrix} A &= \begin{bmatrix} 17 & 4 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 4 & 21 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 0 & 19 \end{bmatrix} A &= \begin{bmatrix} 0 & 19 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 23 & 6 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 18 & 4 \end{bmatrix} A &= \begin{bmatrix} 18 & 4 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 12 & 22 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 21 & 4 \end{bmatrix} A &= \begin{bmatrix} 21 & 4 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 10 & 25 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 13 & 0 \end{bmatrix} A &= \begin{bmatrix} 13 & 0 \end{bmatrix} \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix} = \begin{bmatrix} 0 & 13 \end{bmatrix} \pmod{26} \end{aligned}$$

Converting the ciphertext numbers back into letters yields the following.

6	5	2	11	4	21	23	6	12	22	10	25	0	13
G	F	C	L	E	V	X	G	M	W	K	Z	A	N

Thus, the ciphertext is GFCLE VXGMW KZAN.

Now let A be a 2×2 matrix for which $A^{-1} \pmod{26}$ exists, and suppose we wish to decrypt a ciphertext that was formed using a Hill cipher with the encryption matrix A . We begin by converting the list of letters in the ciphertext into a list of numbers, and then grouping these numbers in order into row matrices of size 1×2 . Suppose the ciphertext is of length n (which would already be a multiple of 2, or, more generally, a multiple of the number of rows in A). For the list of ciphertext numbers $y_1, y_2, y_3, \dots, y_n$, the corresponding list of plaintext numbers $x_1, x_2, x_3, \dots, x_n$ is found by forming the following matrix products.

$$\begin{aligned} \begin{bmatrix} x_1 & x_2 \end{bmatrix} &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} A^{-1} \pmod{26} \\ \begin{bmatrix} x_3 & x_4 \end{bmatrix} &= \begin{bmatrix} y_3 & y_4 \end{bmatrix} A^{-1} \pmod{26} \\ \begin{bmatrix} x_5 & x_6 \end{bmatrix} &= \begin{bmatrix} y_5 & y_6 \end{bmatrix} A^{-1} \pmod{26} \\ &\vdots \\ \begin{bmatrix} x_{n-1} & x_n \end{bmatrix} &= \begin{bmatrix} y_{n-1} & y_n \end{bmatrix} A^{-1} \pmod{26} \end{aligned}$$

The plaintext numbers $x_1, x_2, x_3, \dots, x_n$ are then converted back into letters to give the final plaintext. This shows why the encryption matrix A in a Hill cipher must be chosen so that $A^{-1} \pmod{26}$ exists, since for a ciphertext formed using a Hill cipher with encryption calculation $\mathbf{y} = \mathbf{x}A \pmod{26}$, the corresponding decryption calculation is $\mathbf{x} = \mathbf{y}A^{-1} \pmod{26}$.

With Hill ciphers, all of our calculations will be done with modulo 26 arithmetic. For convenience, each $\det(A) \pmod{26}$ for which $(\det(A))^{-1} \pmod{26}$ exists is given along with $(\det(A))^{-1} \pmod{26}$ in Table 1.

$\det(A)$	1	3	5	7	9	11	15	17	19	21	23	25
$(\det(A))^{-1}$	1	9	21	15	3	19	7	23	11	5	17	25

Table 1: Corresponding values of $\det(A)$ and $(\det(A))^{-1} \pmod{26}$.

To clarify, for

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

with entries in $\mathbb{Z}_{26} = \{0, 1, 2, 3, \dots, 25\}$, $A^{-1} \pmod{26}$ will exist if and only if $\det(A)$ and 26 are relatively prime, that is, if $\gcd(\det(A), 26) = 1$. The values of $\det(A) \pmod{26}$ with $\gcd(\det(A), 26) = 1$ are listed in the first row in Table 1. When $A^{-1} \pmod{26}$ exists, it will be given by the following formula.

$$A^{-1} \pmod{26} = (\det(A))^{-1} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \pmod{26} \quad (1)$$

For this formula, all possible values of $(\det(A))^{-1} \pmod{26}$ are listed in the second row in Table 1, shown in correspondence with the values of $\det(A) \pmod{26}$ listed in the first row.

Consider the encryption matrix $A = \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix}$. Note first that we can find $\det(A) \pmod{26}$ as follows.

$$\det(A) = 8 \cdot 14 - 1 \cdot 19 = 93 = 15 \pmod{26}$$

Since $\gcd(15, 26) = 1$, then we know that $A^{-1} \pmod{26}$ exists. Also, since $\det(A) = 15 \pmod{26}$, Table 1 gives $(\det(A))^{-1} = 7 \pmod{26}$, and we can find $A^{-1} \pmod{26}$ using (1) as follows.

$$\begin{aligned}
A^{-1} \bmod 26 &= 15^{-1} \begin{bmatrix} 14 & -1 \\ -19 & 8 \end{bmatrix} \bmod 26 \\
&= 7 \begin{bmatrix} 14 & -1 \\ -19 & 8 \end{bmatrix} \bmod 26 \\
&= \begin{bmatrix} 98 & -7 \\ -133 & 56 \end{bmatrix} \bmod 26 \\
&= \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix}
\end{aligned}$$

Now, consider the ciphertext OENGD ZHCXG GE, which was formed using a Hill cipher with the encryption matrix $A = \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix}$, whose inverse matrix we just found to be $A^{-1} = \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix}$. Converting the ciphertext letters into numbers yields the following.

$$\begin{array}{cccccccccccc}
O & E & N & G & D & Z & H & C & X & G & G & E \\
14 & 4 & 13 & 6 & 3 & 25 & 7 & 2 & 23 & 6 & 6 & 4
\end{array}$$

To decrypt the ciphertext, we form the following matrix products.

$$\begin{aligned}
[14 \ 4]A^{-1} &= [14 \ 4] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [8 \ 22] \bmod 26 \\
[13 \ 6]A^{-1} &= [13 \ 6] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [8 \ 11] \bmod 26 \\
[3 \ 25]A^{-1} &= [3 \ 25] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [11 \ 1] \bmod 26 \\
[7 \ 2]A^{-1} &= [7 \ 2] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [4 \ 11] \bmod 26 \\
[23 \ 6]A^{-1} &= [23 \ 6] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [0 \ 19] \bmod 26 \\
[6 \ 4]A^{-1} &= [6 \ 4] \begin{bmatrix} 20 & 19 \\ 23 & 4 \end{bmatrix} = [4 \ 0] \bmod 26
\end{aligned}$$

Converting the plaintext numbers back into letters yields the following.

$$\begin{array}{cccccccccccc}
8 & 22 & 8 & 11 & 11 & 1 & 4 & 11 & 0 & 19 & 4 & 0 \\
I & W & I & L & L & B & E & L & A & T & E & A
\end{array}$$

Thus, the plaintext is I WILL BE LATE.

Each of the previous two examples used a Hill cipher in which the encryption matrix A was of size 2×2 . It is not required, of course, that A be of size 2×2 , but rather just that A be square (and that $A^{-1} \bmod 26$ exist). However, as the size of the key matrix for encrypting messages with a Hill cipher increases, the process becomes more computationally intensive. The effect of this can be lessened, of course, through the use of technology.

The fact that Hill ciphers encrypt plaintext numbers in groups rather than one at a time can be viewed as a negative feature since a transcription error in a single ciphertext letter usually leads to more than one error in the decrypted message. However, this negative aspect is far outweighed by the additional security gained by encrypting plaintext numbers in groups. In Section 3, we will consider the additional security that this brings.

We will now demonstrate how a Maplet² can be used to encrypt and decrypt messages with a Hill cipher. The source code for both of the Maplets demonstrated in this paper, as well as directly usable versions of them, can be downloaded at [4]. The code is unique to Maple, but could easily be altered for use with any programming language.

For example, consider a Hill cipher with the following encryption matrix A .

$$A = \begin{bmatrix} 8 & 1 \\ 19 & 14 \end{bmatrix}$$

To use this matrix in encrypting the plaintext BE HERE AT SEVEN, we will use a Maplet entitled **Hill Cipher**, which was written by the authors and designed for this purpose. Figure 1 shows the result of using Maplet to do this, which gives the output GFCLE VXGMW KZAN.

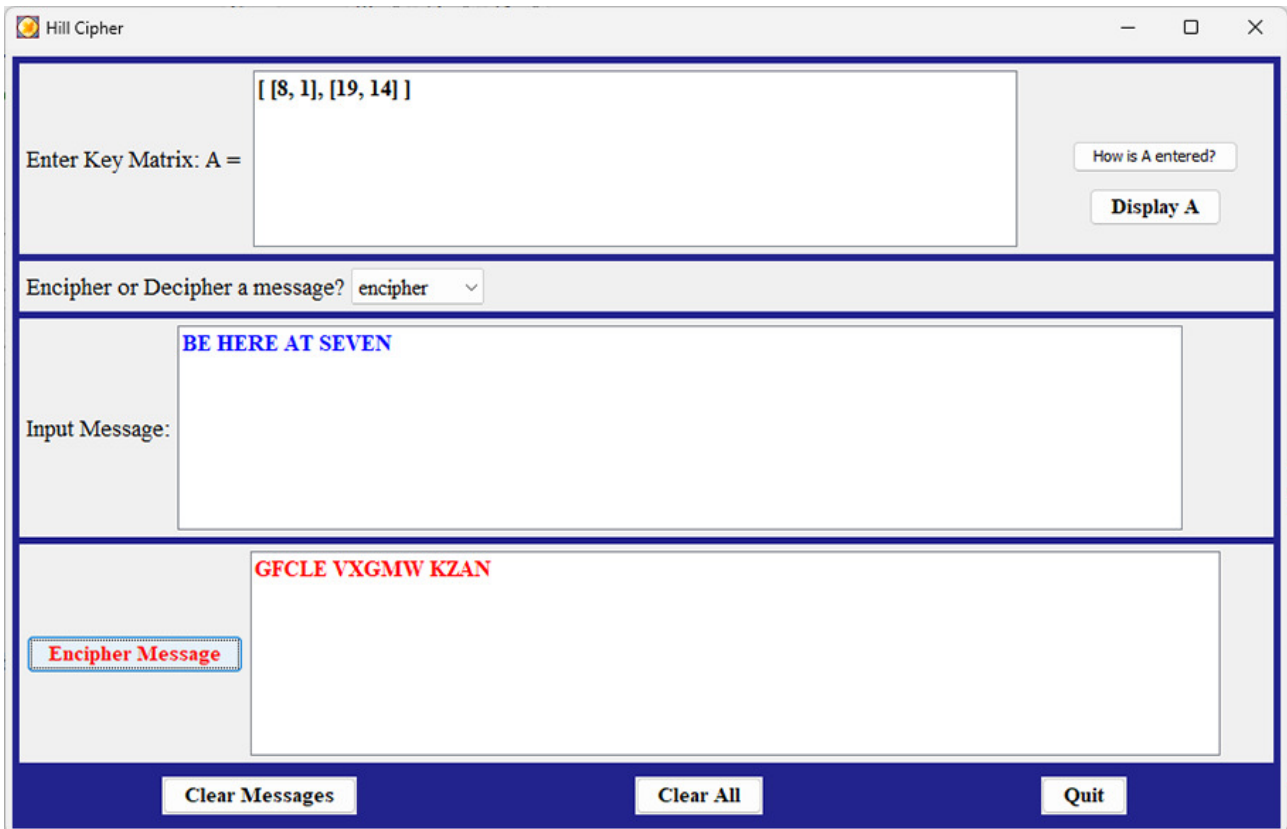


Figure 1: Encrypting a message using a Hill cipher.

As demonstrated in Figure 1, the Maplet requires users to enter the encryption matrix, with the entries in the rows separated by commas, the rows enclosed in brackets and also separated by commas, and a final set of brackets used to enclose the entire matrix. After selecting **encipher** from the

²A Maplet is like an applet, but uses (and requires) the engine of the computer algebra system Maple, and is written using Maple functions and syntax.

Encipher or Decipher a message drop-down menu to indicate that the message is to be encrypted, and typing the plaintext message into the **Input Message** textbox, the message can be encrypted by clicking the **Encipher Message** button, with the resulting ciphertext shown in the box next to this button. If the number of plaintext letters is not a multiple of the size of the key matrix, the Maplet will attach the letter A to the end of the message, repeated as many times as necessary until the last block of letters is of the same length as the number of rows in the key matrix.

To decrypt a message, users enter the key matrix, select **decipher** from the drop-down menu, type the ciphertext, and click the **Decipher Message** button, with the resulting plaintext shown in the box next to this button. The inverse of the key matrix can also be seen by clicking the **Display Inverse Matrix** button. Figure 2 shows the result of using the Maplet to decrypt the ciphertext OENGD ZHCXG GE with the key matrix A that had been used previously, with resulting plaintext I WILL BE LATE.

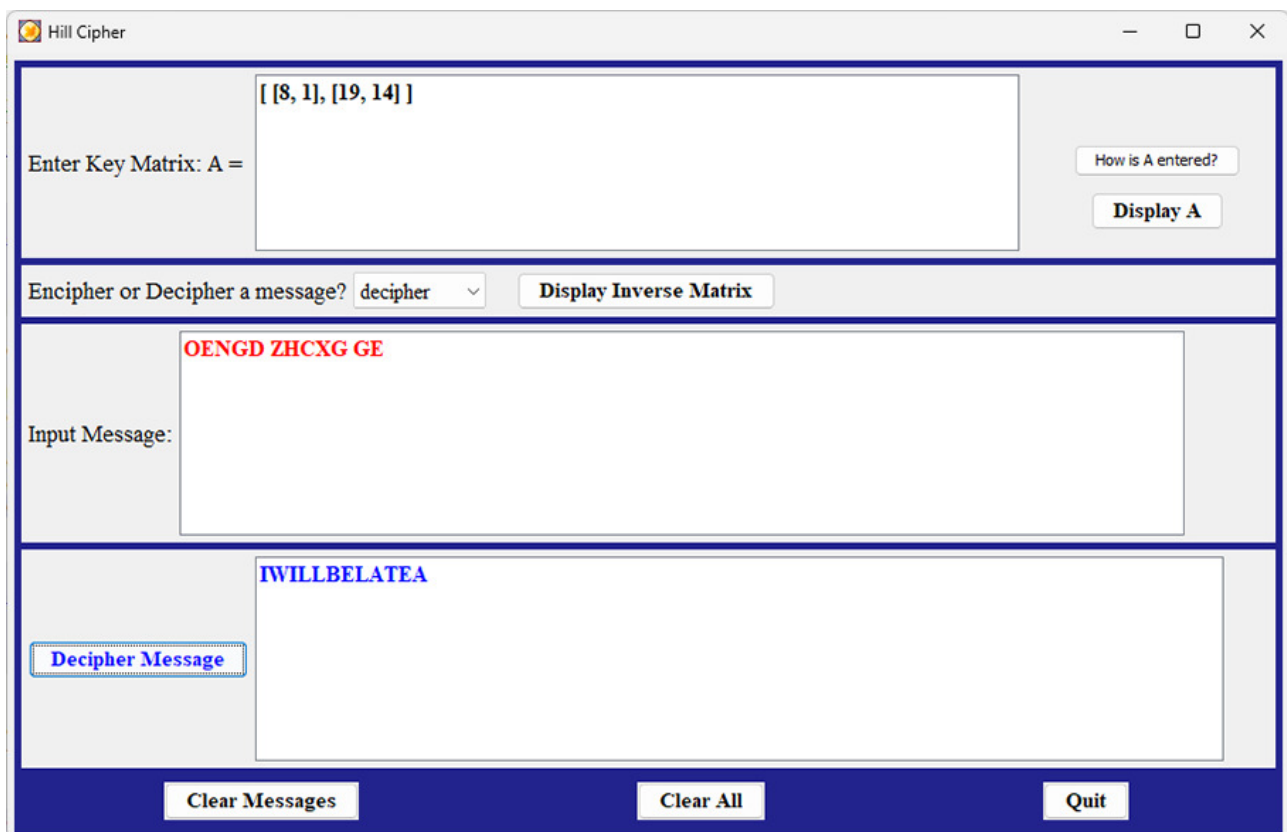


Figure 2: Decrypting a message using a Hill cipher.

One more important thing to consider is that in order for a Hill cipher to be able to decrypt messages uniquely, it is necessary for the key matrix to be invertible. If a key matrix is entered into the **Hill Cipher** Maplet that is not invertible, the Maplet will indicate this and not complete the encryption. To demonstrate this, suppose that the matrix

$$A = \begin{bmatrix} 8 & 2 \\ 19 & 14 \end{bmatrix}$$

is attempted to be used as the key matrix in the encryption of the message BE HERE AT SEVEN. Figure 3 shows that this key matrix is not invertible, and the Maplet does not complete the encryption.

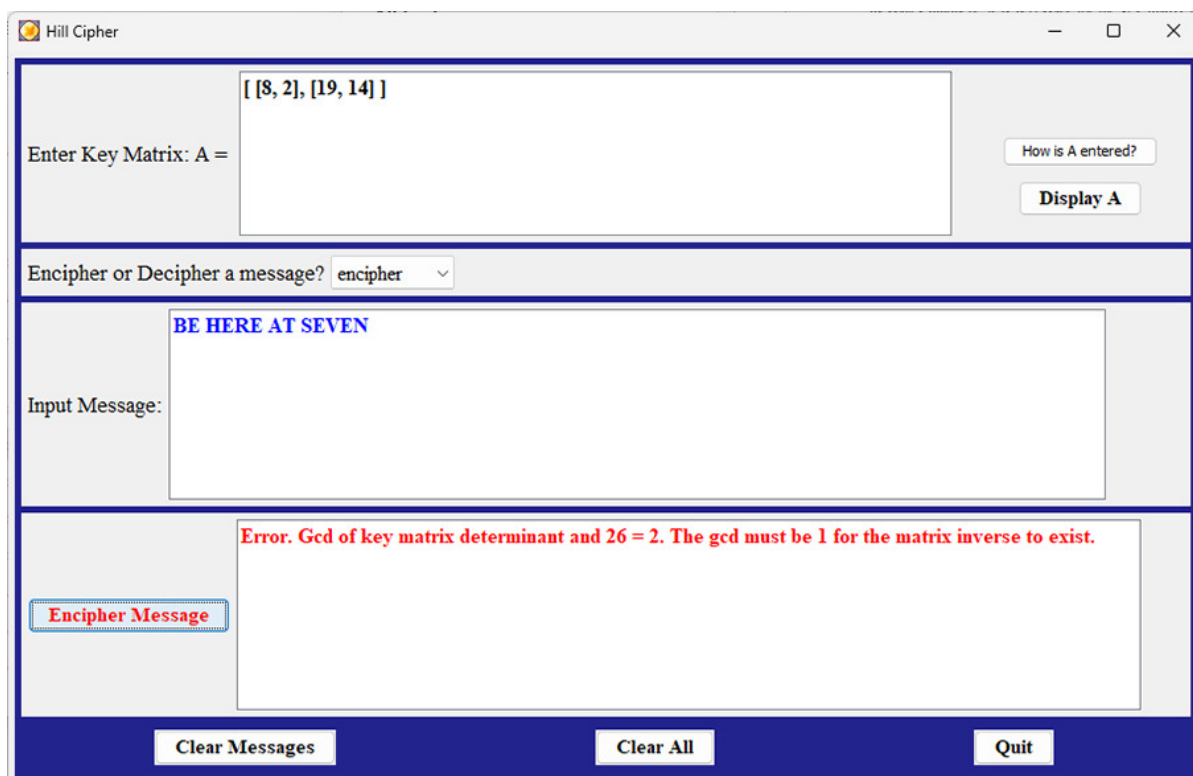


Figure 3: Trying to encrypt a message using a non-invertible key matrix.

3 Hill System Cryptanalysis

Affine ciphers of the form $y = (xa + b) \pmod{26}$ are just substitution ciphers which can be broken fairly easily using frequency analysis, by matching the most frequently occurring letters in the ciphertext with the most frequently occurring letters in the language in which the plaintext was written (which include, in order of frequency in ordinary English, E, T, A, O, I, and N). Correct matches between plaintext and ciphertext letters give a simple system of equations which can be solved and tested to see if the ciphertext decrypts as legible plaintext. We give a detailed description of this in [3].

In Hill ciphers though, plaintext messages are encrypted in groups, which overcomes the insecurity of encrypting messages with substitution ciphers. As noted previously, the fact that Hill ciphers encrypt plaintext numbers in groups rather than one at a time can be viewed as a negative feature since a transcription error in a single ciphertext letter usually leads to more than one error in the decrypted message. However, this negative aspect is far outweighed by the additional security gained by Hill ciphers not being substitution ciphers, which also occurs because plaintext numbers are encrypted in groups rather than one at a time.

Put more plainly, note that while in affine ciphers with encryption formula $y = xa \pmod{26}$ from which Hill ciphers are a generalization, there are only 26 possible plaintext numbers, in Hill ciphers with 2×2 encryption matrices there are $26^2 = 676$ possible plaintext row matrices. Similarly, in Hill ciphers with 3×3 encryption matrices there are $26^3 = 17,576$ possible plaintext row matrices, and in Hill ciphers with 4×4 encryption matrices there are $26^4 = 456,976$ possible plaintext row matrices. Since the encryption matrix A for a Hill cipher can be of any size (provided $A^{-1} \pmod{26}$ exists), and larger encryption matrices allow for more possible plaintext row matrices, the security of Hill ciphers grows as the size of the encryption matrix increases. More precisely, in a Hill cipher with an $n \times n$

encryption matrix, the number of possible plaintext row matrices is 26^n , a number that grows, and grows very quickly, as n increases.

From another perspective, note that in affine ciphers with encryption formula $y = xa \pmod{26}$, there are only 12 possible values of a (the same values corresponding to $\det(A)$ in Table 1), yielding 11 possible keys for the cipher (assuming $a = 1$ is not used). A brute force attack on the cipher could be done by simply trying to decrypt the ciphertext assuming each of these 11 possible keys, stopping when the correct plaintext is revealed. However, a brute force attack on a Hill cipher with a 2×2 encryption matrix would require trying to decrypt the ciphertext assuming a much larger number of possible keys. The number of possible 2×2 matrices with entries in \mathbb{Z}_{26} is $26^4 = 456,976$, and while many of these matrices would not have an inverse modulo 26 and thus not be a valid key matrix for a Hill cipher, it may not be known whether a particular matrix has an inverse modulo 26 until the matrix is at least formed. Thus, a brute force attack on a Hill cipher with a 2×2 key matrix would require some level of testing with up to a maximum of near 456,976 matrices. Similarly, a brute force attack on a Hill cipher with a 3×3 key matrix would require some level of testing with up to a maximum of near $26^9 = 5,429,503,678,976$ matrices. So even for relatively small key matrices, Hill ciphers are much more resistant to a brute force attack than substitution ciphers. More importantly, Hill ciphers can be constructed with any desired level of security by using a key matrix that is sufficiently large.

However, just like with singular characters within languages, two-letter combinations, or *bigrams*, sometimes occur with predictable frequencies. Table 2 shows the relative frequencies of the most commonly occurring bigrams in the ordinary English language.

Bigram	Frequency	Bigram	Frequency
TH	0.0271	AN	0.0161
HE	0.0233	RE	0.0141
IN	0.0203	ES	0.0132
ER	0.0178	ON	0.0132

Table 2: Bigram frequencies in ordinary English.

Letter frequency tables also exist for three-letter combinations, or *trigrams*, and larger combinations, which can be found in [2].

To demonstrate how bigram frequencies can be used to determine a Hill cipher key matrix, suppose we have a ciphertext that was formed using a 2×2 key matrix, and suppose WT and VU are the two most common bigrams in the ciphertext, respectively. Assuming these two bigrams in the ciphertext correspond to the two most frequently occurring bigrams in ordinary English, TH and HE, respectively, we would obtain the following matrix equations.

$$\begin{aligned} \begin{bmatrix} 19 & 7 \end{bmatrix} A &= \begin{bmatrix} 22 & 19 \end{bmatrix} \pmod{26} \\ \begin{bmatrix} 7 & 4 \end{bmatrix} A &= \begin{bmatrix} 21 & 20 \end{bmatrix} \pmod{26} \end{aligned}$$

Equivalently, these two matrix equations can be expressed as the following single matrix equation.

$$\begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix} A = \begin{bmatrix} 21 & 20 \\ 22 & 19 \end{bmatrix} \pmod{26}$$

Using the inverse formula given by (1) with

$$\det \left(\begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix} \right) = 19 \cdot 4 - 7 \cdot 7 = 27 = 1 \pmod{26},$$

we can find the inverse matrix as follows.

$$\begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix}^{-1} = 1^{-1} \begin{bmatrix} 4 & -7 \\ -7 & 19 \end{bmatrix} = \begin{bmatrix} 4 & 19 \\ 19 & 19 \end{bmatrix} \pmod{26}$$

The key matrix A can then be then determined as follows.

$$\begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix} A = \begin{bmatrix} 19 & 7 \\ 7 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 21 & 20 \\ 22 & 19 \end{bmatrix} \pmod{26}$$

$$IA = \begin{bmatrix} 4 & 19 \\ 19 & 19 \end{bmatrix} \begin{bmatrix} 21 & 20 \\ 22 & 19 \end{bmatrix} \pmod{26}$$

$$A = \begin{bmatrix} 487 & 456 \\ 817 & 741 \end{bmatrix} \pmod{26}$$

$$A = \begin{bmatrix} 19 & 14 \\ 11 & 13 \end{bmatrix}$$

To illustrate this cryptanalysis process with an actual message, we can use the **Hill Frequency Breaker** Maplet, which was written by the authors and designed for this purpose. To include an element of randomness, we tested the Maplet using a 200-word sample of ordinary English produced by the AI tool ChatGPT. Figure 4 shows this Maplet displayed with some initial data already entered.

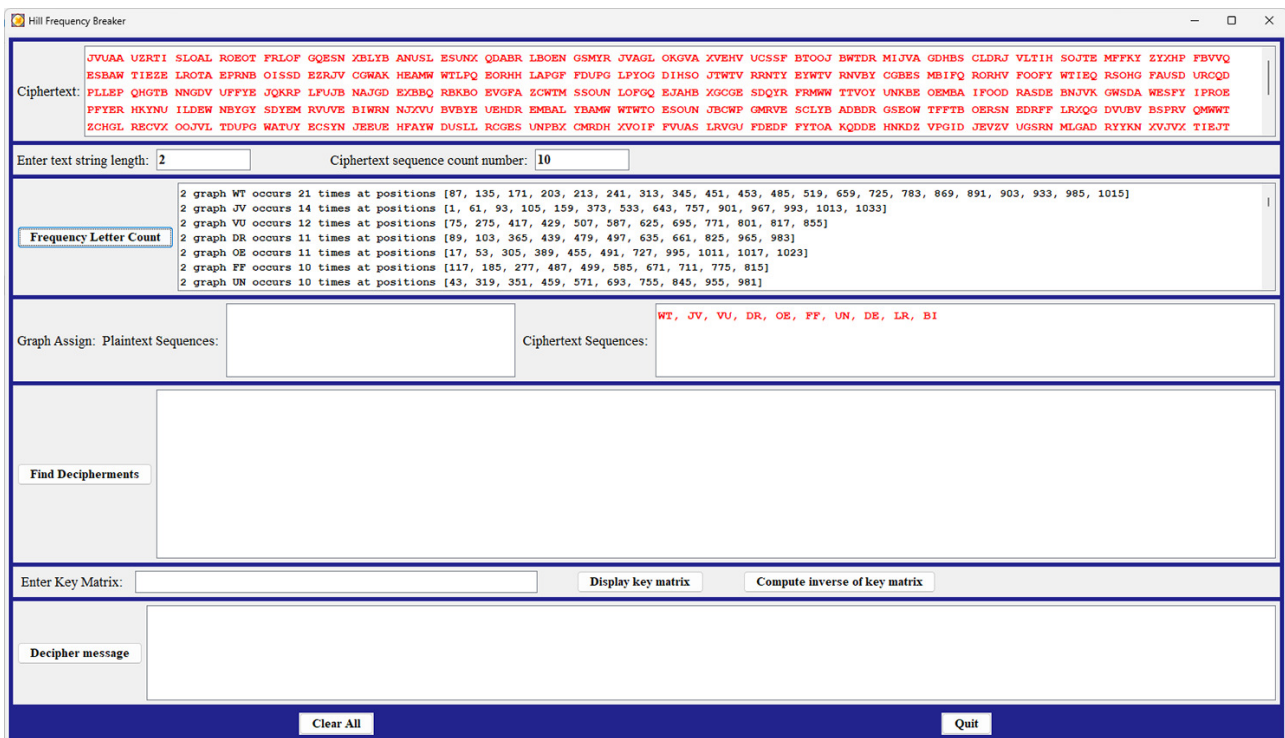


Figure 4: Hill frequency breaker Maplet with initial data entered.

More specifically, we used a 200-word plaintext of random English created by ChatGPT, and encrypted this message using a Hill cipher with a 2×2 key matrix to obtain the full ciphertext JVUAA

UZRTI SLOAL ROEOT FRLOF GQESN XBLYB ANUSL ESUNX QDABR LBOEN GSMYR JVAGL OKGVA XVEHV UCSSF BTOOJ BWTDR MIJVA GDHBS CLDRJ VLTIH SOJTE MFFKY ZYXHP FBVVQ ESBW TIEZE LROTA EPRNB OISSD EZRVJ CGWAK HEAMW WTLPG EORHH LAPGF FDUPG LPYOG DIHSO JTWTV RRNTY EYWTV RNVBY CGBES MBIFQ RORHV FOOFY WTIEQ RSOHG FAUSD URCQD PLLEP QHGTB NNGDV UFFYE JQKRP LFUJB NAJGD EXBBQ RBKBO EVGFA ZCWTM SSOUN LOFGQ EJAHB XGCGE SDQYR FRMWW TTVOY UNKBE OEMBA IFOOD RASDE BNJVK GWSDA WESFY IPROE PFYER HKYNU ILDEW NBYGY SDYEM RVUVE BIWRN NJXVU BVBYE UEHDR EMBAL YBAMW WTWTO ESOUN JBCWP GMRVE SCLYB ADBDR GSEOW TFFTB OERSN EDRFF LRXQG DVUBV BSPRV QMWWT ZCHGL RECVX OOJVL TDUPG WATUY ECSYN JEEUE HFAYW DUSLL RCGES UNPBX CMRDH XVOIF FVUAS LRVGU FDEDF FYTOA KQDDE HNKDZ VPGID JEVZV UGSRN MLGAD RYYKN XVJVX TIEJT ROBEV GKGWT DRNNV QDUDE FFLBO ICSBI TEFNL ZGOBI ESUNV UWRRN PGAAK TKIEH FFFNP FBYTA ZRFRW TOEMK DHPQW MFRXO WCBGL YJTNG DEEMU NJVCG BIPET BWESF VUJIF FFAUS JBWTK EXABI BEJRC ETBGD VUBVO SDZTY LYBIF FVUTI SLOAD RBFRN IUBQI OGMXA BGSKU NFAUS NNGDV UEQOI ZGIOL RNgWT BNUMX QYELP FQEYR CWCNRN WTSOJ TLYJS JWTV RNVBY CGCWD FSUKB IEFQ AVELP MWWT CANUE HGTBP TBNXV XHYEU NRCWA WEUFD RJVKG CGONT YDEEM UNDRW TBNDU IPJVO ECSSF LRIFS OVEDF OEJVV TOEUM SLOEZ GWOVG SRJVF NLBY ECES. This ciphertext is entered in Figure 4, along with the text string length (i.e., the number of rows in the key matrix). When the **Frequency Letter Count** button is clicked, the bigram frequencies that occur two or more times are shown in the box to the right with their positions in the ciphertext. Also, the most frequent bigrams, specifically those occurring at least as many times as the number entered in the **Ciphertext sequence count number** textbox, are shown in the **Ciphertext Sequences** textbox.

Next, we enter the **Plaintext Sequences** that we would like to test with the corresponding ciphertext sequences to determine a possible 2×2 key matrix. Since the bigrams TH and HE occur most frequently according to Table 2, we enter those separated by commas and then click the **Find Decipherments** button to look for possible key matrices. The result is shown in Figure 5.

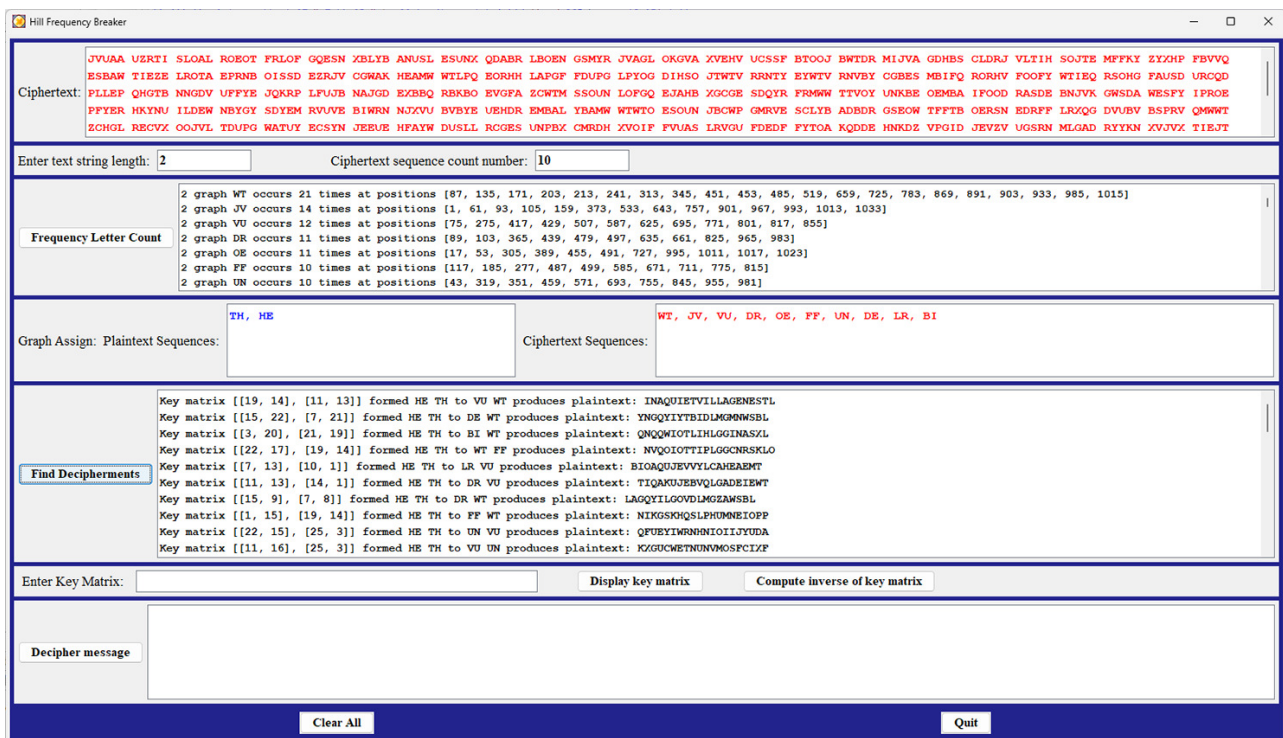


Figure 5: Searching for possible key matrices.

The Maplet specifically attempts to match pairs of plaintext/ciphertext bigrams to produce possible key matrices. For each of the possible key matrices produced, the first twenty characters of the resulting plaintext are given, and we would hope that one is the beginning of a readable English message. To assist in the search, the answers are sorted to give potential plaintexts with the largest number of letters that we normally find in ordinary English sentences, which include, in order, E, T, A, O, I, and N. Examining the output, the plaintext IN A QUIET VILLAGE NESTL is produced by assigning the plaintext bigrams HE and TH to the ciphertext bigrams VU and WT, respectively. The resulting key matrix produced by these assignments is given by

$$\begin{bmatrix} 19 & 14 \\ 11 & 13 \end{bmatrix}.$$

The entire plaintext can be found by entering the key matrix in the **Enter Key Matrix** textbox and clicking the **Decipher message** button. The result is shown in Figure 6.

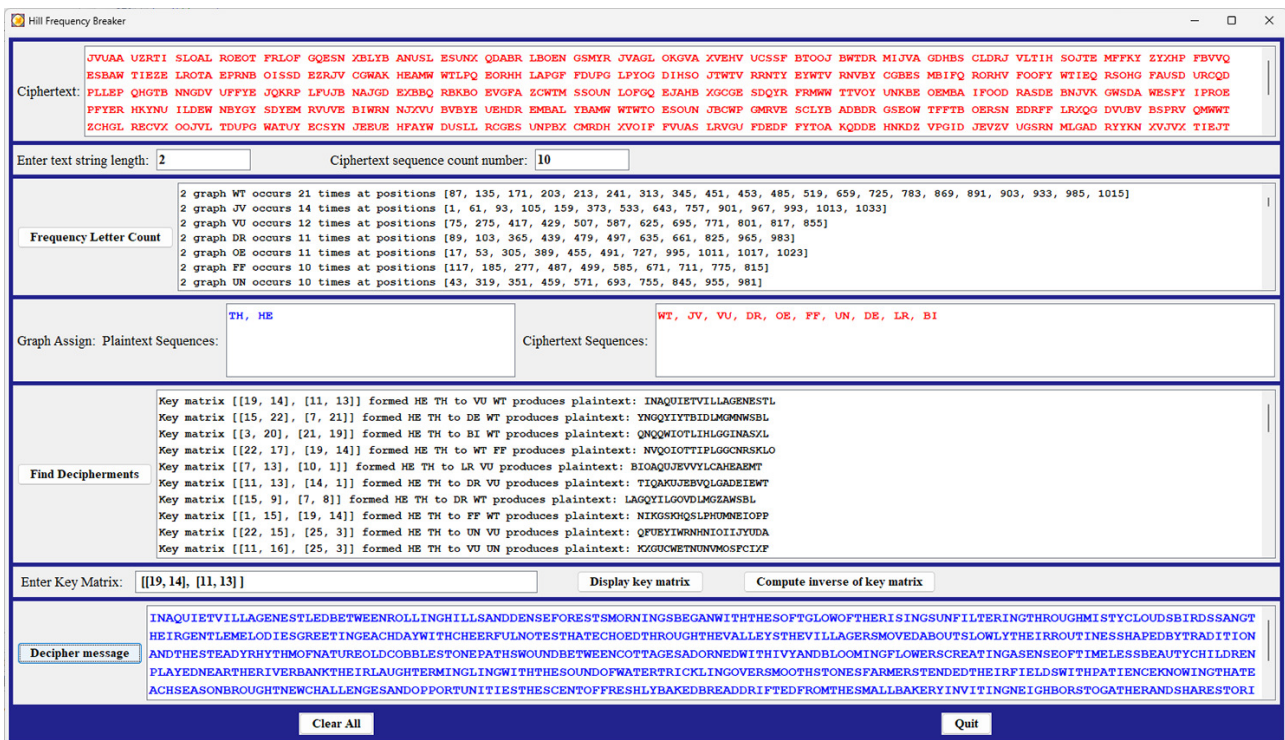


Figure 6: Decrypting a message that was encrypted using a 2×2 key matrix.

The resulting plaintext is IN A QUIET VILLAGE NESTLED BETWEEN ROLLING HILLS AND DENSE FORESTS MORNINGS BEGAN WITH THE SOFT GLOW OF THE RISING SUN FILTERING THROUGH MISTY CLOUDS BIRDS SANG THE GENTLE MELODIES GREETING EACH DAY WITH CHEERFUL NOTES THAT ECHOED THROUGH THE VALLEYS THE VILLAGERS MOVED ABOUT SLOWLY THEIR ROUTINES SHAPED BY TRADITION AND THE STEADY RHYTHM OF NATURE OLD COBBLESTONE PATHS WOUND BETWEEN COTTAGES ADORNED WITH IVY AND BLOOMING FLOWERS CREATING A SENSE OF TIMELESS BEAUTY CHILDREN PLAYED NEAR THE RIVERBANK THEIR LAUGHTER MINGLING WITH THE SOUND OF WATER TRICKLING OVER SMOOTH STONES FARMERS TENDED THEIR FIELDS WITH PATIENCE KNOWING THAT EACH SEASON BROUGHT NEW CHALLENGES AND OPPORTUNITIES THE SCENT OF FRESHLY BAKED BREAD DRIFTED FROM THE SMALL BAKERY INVITING NEIGHBORS TO GATHER AND SHARE STORIES OVER

WARM LOAVES AND HERBAL TEA AT NIGHT STARS BLANKETED THE SKY UNDIMMED BY CITY LIGHTS REMINDING EVERYONE OF THE VASTNESS OF THE UNIVERSE BEYOND THEIR SIMPLE LIVES THE VILLAGERS VALUED COMMUNITY KINDNESS AND THE QUIET MOMENTS THAT MADE EACH DAY SPECIAL THOUGH LIFE IN THE VILLAGE WAS HUMBLE IT WAS RICH WITH CONNECTION GRATITUDE AND PEACE OFFERING A GENTLE REMINDER THAT HAPPINESS OFTEN FLOURISHES IN THE SMALLEST MOST ORDINARY PLACES.

The Maplet will work on any large ciphertext, including those encrypted with key matrices of size $n \times n$ with $n > 2$. However, through different trials when testing the limits of decoding a 2×2 cipher with the code, we found that ciphertexts typically need to be at least 200 words in length. For key matrices of size $n \times n$ with $n > 3$ though, the ciphertexts typically needs to be much larger. Figure 7 gives an example of a recovered plaintext of approximate length 450 words that was encrypted using a Hill cipher with the 3×3 key matrix

$$\begin{bmatrix} 11 & 6 & 8 \\ 0 & 3 & 14 \\ 24 & 0 & 9 \end{bmatrix}.$$

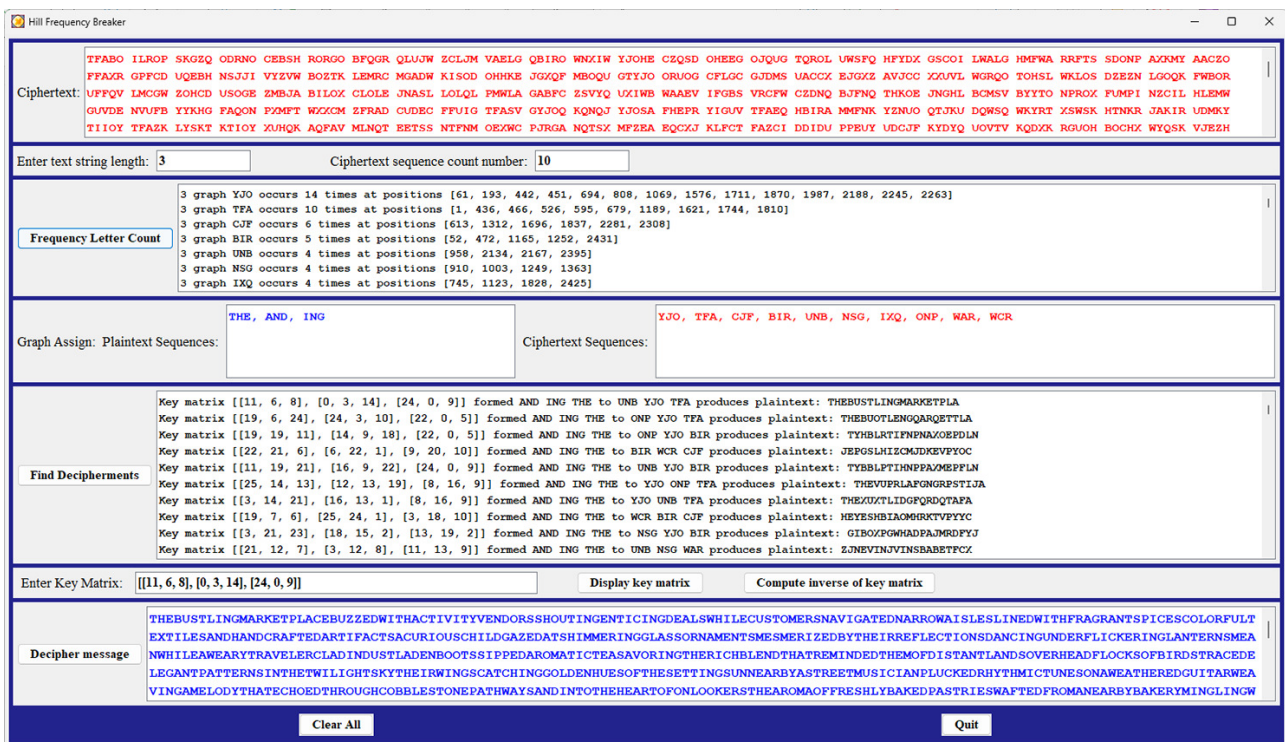


Figure 7: Decrypting a message that was encrypted using a 3×3 key matrix.

The key matrix in this example was found by solving the matrix equation created by assigning three of the most common English trigrams, THE, AND, and ING, to the three-letter ciphertext sequences TFA, YJO, and UNB, respectively.

Without a large amount of ciphertext to work with, the process of breaking a message that was encrypted using a Hill cipher is much more difficult. However, if a small part of the plaintext, called a *crib*, is somehow known, it can be much easier to break a Hill cipher for larger key matrices and shorter ciphertexts. We discuss this process in extensive detail in [3].

4 Conclusion

In this paper, we examined how Hill ciphers can be broken through frequency analysis of letter combinations. This provides a method for illustrating linear algebra concepts in a practical setting.

Other cipher systems, including some modern ones such as the Advanced Encryption Standard, use matrices and linear algebra concepts in their implementation. We provide more information about these and other cipher systems in [3].

References

- [1] Lester Hill. Cryptography in an algebraic alphabet. *American Mathematical Monthly*, 36:306–312, 1929
- [2] English letter frequencies. *Practical Cryptography*, 2025. Available at: <http://www.practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/english-letter-frequencies/>
- [3] Richard Klima and Neil Sigmon, *Cryptology, Classical and Modern, Second Edition*, Taylor & Francis, Boca Raton, FL, 2019.
- [4] Neil Sigmon, 2025. Maplet Download Page for Using Bigram Frequencies to Break Hill Ciphers. Available at: <https://sites.radford.edu/~npsigmon/Hillcipher/paper.html>.