

ELEMENTS OF CRYPTANALYSIS

Training Pamphlet No. 3

PREPARED IN THE OFFICE OF THE
CHIEF SIGNAL OFFICER

May, 1923

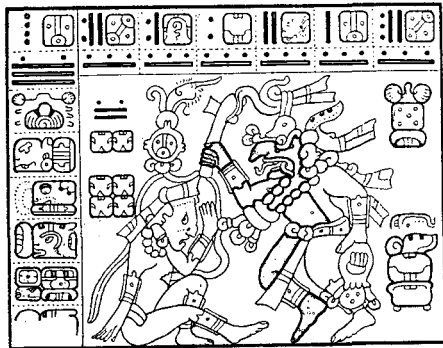


Box 25

WASHINGTON
GOVERNMENT PRINTING OFFICE
1924

William F. Friedman
Washington
1924 - 3

Umol-huun tah-tiyal
William Frederick
yetel
Elizebeth Smith Friedman



Lay ca-buunil kubenbil tech same.
This our book we entrusted you a while-ago.
Ti manaan apaclam-tz'a lo toon
It not-being you-return-give it us,
Epabal ca-baat tumen ab-men.
Is-being-sharpened our-axe by the expert.

213

~~FOR OFFICIAL USE ONLY~~

ELEMENTS OF CRYPTANALYSIS

Training Pamphlet No. 3

PREPARED IN THE OFFICE OF THE
CHIEF SIGNAL OFFICER

May, 1923



WASHINGTON
GOVERNMENT PRINTING OFFICE
1924

Certificate: By direction of the Secretary of War, the matter contained herein is published as administrative information and is required for the proper transaction of the public business.

WAR DEPARTMENT
Document No. 1117
Office of The Adjutant General

ii

The follo
Training Pa
information
[A. G. 062.1

By ORDE

OFFICIAL
ROI

This c
of the

WAR DEPARTMENT,
WASHINGTON, *May 12, 1923.*

The following publication, entitled "Elements of Cryptanalysis," Training Pamphlet No. 3 (for official use only), is published for the information and guidance of all concerned.

[A. G. 062.12 (4-21-23).]

BY ORDER OF THE SECRETARY OF WAR:

JOHN J. PERSHING,
*General of the Armies,
Chief of Staff.*

OFFICIAL:

ROBERT C. DAVIS,
The Adjutant General.

III

30 April 1959

This document is declassified by authority
of the Director, National Security Agency.

Paul S. Willard
Paul S. Willard
Colonel, AGC
Adjutant General

PREFACE.

The material contained in this pamphlet forms the basis of a course in Military Codes and Ciphers given at The Signal School, Camp Alfred Vail, N. J., by Capt. W. F. Friedman, Sig. O. R. C., cryptanalyst in the Office of the Chief Signal Officer. This course is intended to give a brief exposition of the general subject of military cryptography, to show how and why certain types of cryptograms are solved so readily, and to point out and exemplify the various rules and precautions that should be observed in order to maintain the secrecy of our own communications. No attempt is made in this course to give instruction in the analysis of the more complex types of cryptograms, inasmuch as a longer period of instruction and study would be necessary for such advanced work than is usually available, but it is believed that the ground covered and the material contained herein afford a firm general foundation for the further training of students in the application of the principles of cryptanalysis to the more complicated types of cryptograms such as would be encountered in any future emergency.

TABLE OF CONTENTS.

PART 1. THE ANALYSIS OF CIPHERS.		Paragraphs.
Section I.	Preliminary definitions and explanations.....	1-6
II.	The elements of an alphabetical language.....	7-12
III.	Kinds of alphabets.....	13-15
IV.	Monoalphabet and polyalphabet substitution ciphers.....	16-17
V.	Solution of monoalphabet substitution ciphers using standard alphabets.....	18-19
VI.	Solution of monoalphabet substitution ciphers using mixed alphabets.....	20-27
VII.	Remarks on the solution of monoalphabet ciphers.....	28-29
VIII.	Introductory remarks on polyalphabet substitution ciphers.....	30-34
IX.	Solution of periodic polyalphabet ciphers using standard alphabets.....	35-40
X.	Solution of periodic polyalphabet ciphers using mixed alphabets.	
	Case I. The plain component is the normal sequence.....	41-46
XI.	Solution of periodic polyalphabet ciphers using mixed alphabets.	
	Case II. Both components are mixed sequences.....	47-48
XII.	Solution of progressive polyalphabet ciphers.....	49-51
XIII.	Solution of nonperiodic polyalphabet ciphers.....	52-56
XIV.	Miscellaneous substitution methods.....	57-61
XV.	Transposition ciphers—standard methods of solution.....	62-67
XVI.	Transposition ciphers—special methods of solution.....	68-71
XVII.	Concluding remarks on military ciphers.....	72-75
XVIII.	Frequency data for other languages.....	76-83
PART 2. THE ANALYSIS OF CODES.		
XIX.	Code systems and code books.....	84-91
XX.	Analysis of codes.....	92-95
XXI.	Solution by detailed analysis or first principles.....	96-101
XXII.	Solution by comparison or analogy.....	102-103
XXIII.	Miscellaneous considerations.....	104-107
XXIV.	Bibliography.....	108
Appendix.	Problems given during course at Camp Vail.....(Page)..	145

PART 1.—THE ANALYSIS OF CIPHERS.

SECTION I.

PRELIMINARY DEFINITIONS AND EXPLANATIONS.

	Paragraph.
Cryptography.....	1
Types of cryptograms.....	2
Enciphering and encoding.....	3
Enciphered code.....	4
Deciphering and decoding.....	5
Cryptanalytic.....	6

1. Cryptography.—Cryptography is the science which embraces all the methods and devices whereby an intelligible, written message may be converted into an unintelligible or secret form. The words of the original or intelligible message constitute the **PLAIN TEXT**; the characters of the unintelligible or secret form of the message constitute the **CRYPTOGRAM**.

2. Types of cryptograms.—Cryptograms are of two more or less distinct types: (1) cipher messages, (2) code messages.

A **CIPHER** message is a cryptogram which has been produced by applying a method of cryptography to the *individual letters* of the plain text, taken either singly or in groups of constant length.¹ Practically every cipher message is the result of the joint application of two elements: (1) A **GENERAL SYSTEM** or method of treatment which, once agreed upon, is invariable or unchanging in nature, and (2) a **SPECIFIC KEY**, which is variable or changeable at the will of the correspondents, and controls or determines the exact steps to be followed under the general system.

A **CODE** message is a cryptogram which has been produced by the use of a *code book* consisting of arbitrary combinations of letters or figures to be substituted for the entire words, phrases, and sentences, and sometimes individual letters or syllables, of the plain text.

Although a code system is in reality only a highly specialized form of a particular kind of cipher system, the principal distinction between the two types of systems is that in ciphers one usually deals with the individual letters, or definite groups of letters, taken as units, whereas in codes one deals with entire words, phrases, or even whole sentences taken as units. The differences between the two systems are quite marked.

3. Enciphering and encoding.—The operation or process of converting the plain text into its equivalent cryptographic text is called

¹ The only exception to the general rule as regards the treatment of individual letters is in the case of the now obsolete "Route Cipher" once employed by the Federal Army. Here whole words are treated as units. See p. 60.

ENCIPHERING if a cipher system is involved, or **ENCODING** if a code system is involved. The text of the cryptograms thus produced is called **CIPHER TEXT** or **CODE TEXT**, respectively, as the case may be. To **ENCIPHER** a message is to apply some process of **ENCIPHERMENT** to it, whereby it is converted into a cipher. To **ENCODE** a message is to apply the process of **ENCODEMENT** to it, that is, to replace the elements of the message by the arbitrary combinations of letters or figures, which are termed **CODE GROUPS**, or very often in the case of combinations of letters, **CODE WORDS**.

4. Enciphered code.—Sometimes the code groups of a code message undergo a further process of encipherment, in which case the resulting cryptogram constitutes an **ENCIPHERED CODE MESSAGE**. This is used only under special circumstances where a high degree of secrecy must be maintained, or where the code book employed is not itself kept secret.

5. Deciphering and decoding.—The operation or process of reconverting the text of a cryptogram into its equivalent plain text, when this is accomplished directly, in the case of ciphers, by a knowledge of the system and the key employed, or, in the case of codes, by the possession of the code book used in producing the cryptogram, is called **DECIPHERING** in the case of a cipher message, and **DECODING** in the case of a code message. The plain text thus produced is often called the **DECIPHER** or **DECODE**, respectively, as the case may be. To **DECIPHER** a message is to apply the necessary process of **DECIPHERMENT** to it, whereby it is reconverted into the plain text; to **DECODE** a message is to apply the process of **DECODEMENT** to it, that is, to replace the code groups by the plain-text groups they represent by reference to the code book. It is obvious that the correspondents must possess identical copies of the code book. In the case of an enciphered code message the processes of decipherment and decodement must both be applied, usually in the order given, rarely in the reverse order. For the sake of brevity the term **REDUCTION** is here presented to apply to either decipherment or decodement, or both, where no confusion is likely to arise. To **reduce** a cryptogram is to produce the plain text, whether the cryptogram be a code message, or a cipher message, or an enciphered code message, and it is implied that such reduction is accomplished directly and quickly by means of a knowledge of the system and the key in the case of cipher messages, or the possession of the code book in the case of code messages.

6. Cryptanalytics.—Cryptanalytics is the name recently applied to the science which embraces all the principles, methods, and means employed in the **ANALYSIS** of cryptograms, that is, their reduction or solution without a knowledge of the system or the key, or the

possession of the code book, by a detailed study of the cryptograms themselves. CRYPTANALYSIS is the name applied to the steps performed in the application of the principles of cryptanalytics to cryptograms. A CRYPTANALYST is an expert in the application of the operations or processes in cryptanalysis.

The principal aim in this pamphlet is to subject the more common types of cipher systems to a careful scrutiny, to point out the ease with which certain types are analyzed, or the difficulty with which other types are solved, and thus to demonstrate by inference some of the reasons for the adoption or rejection of certain methods of cryptography by the Signal Corps. It should be understood that this is but an elementary treatise on the subject, and that only the most important phases of the science can be dealt with in detail in these few pages. For further treatment the bibliography given at the end should be consulted.

SECTION II.

THE ELEMENTS OF AN ALPHABETICAL LANGUAGE.

	Paragraph.
Nature of the English language.....	7
Frequency tables.....	8
Substitution ciphers.....	9
Transposition ciphers.....	10
Substitution and transposition ciphers compared.....	11
Method of distinguishing transposition from substitution ciphers.....	12

7. Nature of the English language.—The English language is written by means of a set of 26 simple characters called *letters*, which, taken together and considered as a *sequence* of symbols, constitute an *alphabet*. Not all written languages are of this nature. The Chinese language, for example, is written by means of some 40,000 more or less complex characters, each representing one sense of a word, and these words are all monosyllables. The written languages of the majority of other civilized peoples of to-day are, however, alphabetical in construction, so that the principles discussed herein will apply in general to all of them, with such modifications only as are necessitated by the special characteristics of the language concerned. We shall, therefore, only treat of the analysis of messages written in the English language, and leave to the discretion of the individual student who may be interested in further work, practice in the analysis of cryptograms in other languages.

8. Frequency tables.—(a) If a tabulation, called a *frequency table*, is made of the number of letters appearing in a fairly large volume of ordinary English literary text, some interesting facts are disclosed. Figure 1 is an example of such a table compiled by Hitt¹ from a count of 10,000 letters taken from military orders and reports.

¹ HITT, Lieut. Col. Parker. *Manual for the Solution of Military Ciphers*. Army Service Schools Press, Fort Leavenworth, Kans. 1918.

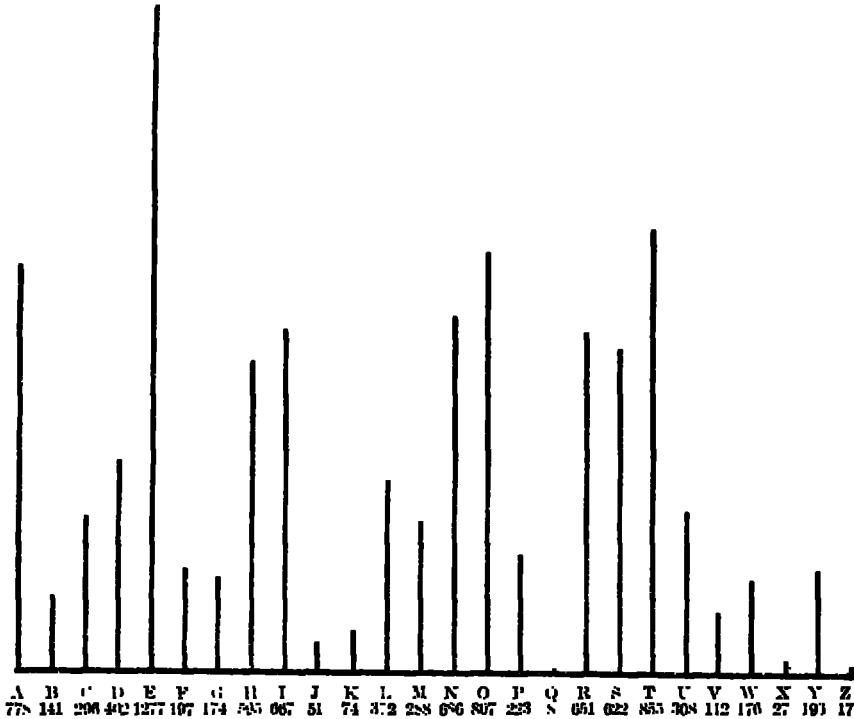


FIGURE 1.—Frequency table for normal English text.

The following are the most important facts disclosed by this table, which has been made in the form of a chart:

- (1) It shows a very irregular appearance due to the fact that some of the letters are used very frequently, others rarely. This gives rise to the presence of what we shall hereafter call *crests* and *troughs*; that is, points of high frequency and low frequency. These crests and troughs occupy more or less definitely fixed positions in the chart, that is, their *spatial relations* are predetermined by circumstances to be discussed later; furthermore, their relative heights and depths, that is, their *linear extensions* are also more or less fixed, as would be found if a similar volume of text be analyzed.
- (2) The most prominent crests are marked by the letters E, T, O, A, N, I, R, S, and H; the most prominent troughs are marked by the letters J, K, Q, X, and Z.
- (3) The following are the proportions of vowels and consonants to the total number of letters:

Vowels A E I O U Y	40.33%	40.33%
High-frequency consonants H N R S T	34.09	} 59.67%
Medium-frequency consonants		
D L C M P F W G B V	23.81	
Low-frequency consonants J K Q X Z	1.77	
Total	100.00	100.00

(4) The relative order of frequency of all the letters is as follows:

E	1277	I	667	L	372	F	197	V	112
T	855	R	651	U	308	Y	196	K	74
O	807	S	622	C	296	W	176	J	51
A	778	H	595	M	288	G	174	X	27
N	686	D	402	P	223	B	141	Z	17
								Q	8

(b) The data given above represent, of course, the relative frequencies found in a fairly large volume of text; the frequencies will vary somewhat with the nature of the text examined. For example, telegraphic text, because it generally omits such words as "the," "that," etc., and employs somewhat longer words, shows slightly different frequencies, as given in the following table based upon a count of 10,000 letters of ordinary telegraphic text, given by Hitt:

E	1319	R	677	H	386	Y	208	V	136
O	844	S	656	U	321	F	205	K	88
A	813	T	634	C	306	G	201	X	51
N	718	D	417	M	273	W	166	J	42
I	711	L	392	P	243	B	149	Q	38
								Z	6

Aside from the frequencies of the letters T and H, these two sets of frequencies are practically the same. If similar counts be made of the letters occurring in other bodies of English text of the same volume as those above, the relative frequencies will be found to be practically identical with them. In short, the frequencies disclosed by such a large count may be considered to be the standard or normal frequencies of the letters in the English language; counts made of smaller volumes of text will tend to approximate these normal frequencies, and the smaller the volume, the lower will be the degree of approximation to the normal, until, in a very short message, the normal proportions may not hold at all. It is advisable that the student fix this fact firmly in mind, for the sooner he realizes the true nature of frequency tables, namely, that they are merely statistical generalizations that will hold only in large volumes of text, and may not even be approximated in small volumes of text, the more rapid will be his progress in cryptanalysis.

(c) It has been found that the *average* length of messages is approximately 200 letters, and that the proportions of vowels, high-frequency, and low-frequency consonants differ by approximately 5 per cent above or below, from the normal proportions as given above. It will be convenient for our purposes to reduce the frequencies for literary and telegraphic text to a basis of 200 letters to facilitate comparison with the average length of cipher messages. The best

way to exhibit these data is in the form of a *graphic frequency table*, shown in Figures 2 and 3.

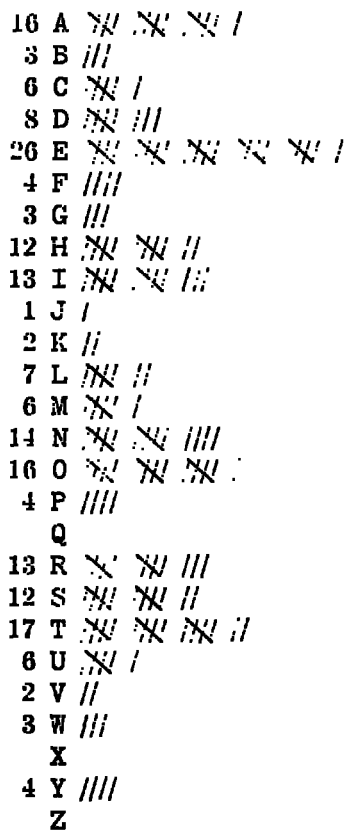


FIGURE 2.—(Basis of 200 letters) Literary text.

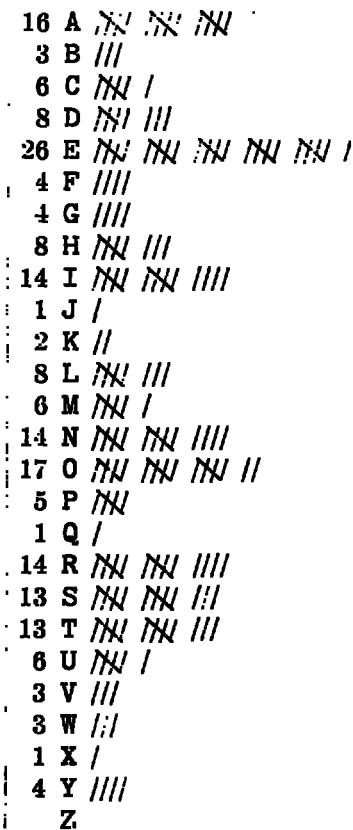


FIGURE 3.—(Basis of 200 letters) Telegraphic text.

9. Substitution ciphers.—The letter E is the most frequently used letter in our language. A little reflection will show that the characteristic of being the most frequently used letter in the English language belongs to the letter E, not because of any special peculiarity inherent in the *symbol* E, but because it is the character which represents the *sound* that happens to be the most frequently used in our language. The reason for this does not concern us; it probably involves the very fundamentals and origin of the language itself. The characteristics of the other letters of the alphabet as regard frequency rest upon the same basis; each one has its characteristic frequency, based upon the frequency of the elementary sound it represents. It is recognized, of course, that some of our letters represent more than one elementary sound, and this, it must be admitted, is one of the defects in our alphabet, but for our purposes we need consider that each letter stands for but one sound. Now, the *spoken* words of a language are merely these combinations and permutations of elementary *sounds*

which have by long usage become recognized and adopted as the representatives of definite objects or ideas; the *written* words of an alphabetical language are merely the combinations and permutations of simple *symbols* which represent the elementary sounds of which the words are composed. The symbols or letters we have to-day are the results of a long period of evolution, for the most part still unknown, or lost in the haze of antiquity. Any other symbols, so long as the sounds which they represent are understood by those concerned, will serve the same purpose. Thus, for example, if we were taught from early childhood that the symbols S, *, and @ represent the sounds "Ay," "Bee," and "See," respectively, the combination @S* would still be pronounced "CAB," and would, of course, have exactly the same meaning as before. Or let us suppose that two persons have agreed to change the sound value of the letters E, F, and G, and after long practice have become accustomed to pronouncing them as "Ay," "Bee," and "See," respectively. They would then see nothing strange in the fact that they pronounce the combination GEF as we do "CAB," but to us and to others not party to their agreements "GEF" constitutes cipher. The combination of sounds called for by this combination of symbols is perfectly intelligible to the two who have adopted the new sound values for those symbols, and therefore pronounce GEF as "CAB," but is unintelligible to us who are reading it on our own and long-established sound value basis, and therefore pronounce it as "GEF." We would say that there is no such word as "GEF," by which we mean merely that this particular combination of sounds has not been adopted by convention to represent a thing or an idea in our language. Thus we see that in order for the written words of a language to be intelligible to all who speak that language, the sound values of the letters or symbols used must be universally understood and agreed upon. The simplest method of establishing and maintaining the equivalency of the sounds and symbols is to give each symbol a more or less definite sound value and then fix the order of the symbols. This is what our alphabet really does, and when it is taught us we learn three things: First, the number of symbols or letters; secondly, their sound values; and, thirdly, their normal order. If we now replace the letters of any plain-text message by other letters whose sound values are different from those of the original letters, we have a type of cryptogram called a *substitution cipher*. We may replace single letters or combinations of letters by various other characters, signs, numbers, or combinations of them, and these would also constitute substitution ciphers. A general definition is therefore as follows:

A substitution cipher is a cryptogram in which the original letters of the plain text, taken either singly or in groups of constant length, have been replaced by other letters, figures, signs, or combinations of them, in accordance with a definite system and key.

10. Transposition ciphers.—If, however, we employ the letters of the alphabet with their conventional sound values in combinations which have not been established by convention as representatives of things or ideas, we again have cipher, but of a different type. For example, if we take the letters B, O, and Y of the word BOY and rearrange them to form the sequence YOB, the combination is unintelligible to us, because the particular sequence of sounds called for by this sequence of symbols has not been established as representing a thing or an idea in our language. It is cipher, not because we have changed the sound value of the letters, but because we have altered the arrangement of the letters, that is, their relative order. This type of cryptogram is called a *transposition cipher* and is defined as follows:

A transposition cipher is a cryptogram in which the original letters of the plain text have merely been rearranged according to a definite system and key.

11. Substitution and transposition ciphers compared.—The fundamental difference between substitution and transposition methods is that in the former the normal or conventional values of the letters of the plain text are changed, without any change in the relative positions of the letters in their original sequences, whereas in the latter only the relative positions of the letters of the plain text in the original sequences are changed, without any change in the conventional values of the letters. Every cipher falls into one of these two principal classes, and since the method of encipherment is radically different in the two cases, the principles involved in the analysis of substitution ciphers are fundamentally different from those of transposition ciphers.

12. Method of distinguishing transposition from substitution ciphers.—The first step in the analysis of a cipher is, therefore, to determine the class to which it belongs. Cryptograms composed of symbols, characters, figures, and the like are patently substitution ciphers, and hence no further examination is necessary to determine the class to which they belong. Cryptograms composed exclusively of letters, and these form the great majority, are subjected to a count of the total number of letters, the number of vowels, high-frequency consonants, and low-frequency consonants. If the percentages found agree within 5 per cent above or below the normal, as given on page 1, the cryptogram is classed as a transposition cipher; if these percentages differ from the normal by more than 5 per cent, then it is classed as a substitution cipher.

The reason why this simple determination is possible should be obvious. Since in a transposition cipher the original letters of the plain text have merely been rearranged, without any change whatsoever in the identity or conventional values of the letters themselves, the numbers of vowels, high-frequency and low-frequency consonants remain exactly the same in the cryptogram as in its equivalent plain text message, and therefore their proportions remain the same. As has been stated in paragraph 10, in messages of average length, these proportions will usually agree within a per

cent, above or below, with the normal proportions for English plain text. Reasoning conversely, when the proportions of vowels, high-frequency and low-frequency consonants in a cryptogram agree within 5 per cent with the normal, it follows that these elements are present in the proper numbers to form plain text, and therefore the cryptogram is to be classed as a transposition cipher. On the other hand, since in a substitution cipher the original letters of the plain text have been substituted among themselves, high-frequency letters being replaced by letters of normally low-frequency, vowels for consonants, and so on, the numbers of vowels, high-frequency and low-frequency consonants in the resulting cryptogram will usually be markedly different from their numbers in its equivalent plain-text message, and therefore their proportions will no longer agree even approximately with the proportions found in normal plain text. Again reasoning conversely, when the proportions of vowels, high-frequency and low-frequency consonants in a cryptogram do not agree within 5 per cent with the normal, it follows that these elements are not present in the proper numbers to form plain text, and therefore the cryptogram is to be classed as a substitution cipher. Usually the proportions in a substitution cipher are very different from the normal. Occasionally one may encounter a cipher which has been written by interchanging letters of approximately similar normal frequencies, vowels for vowels, and consonants for consonants, thus yielding a cryptogram giving external indications of being a transposition cipher, but which is really a substitution cipher. A close study of the message along the lines to be discussed later will soon show the futility of so simple a subterfuge.

We shall proceed first to study the most common types of substitution methods, after which transposition methods will be taken up, and then a brief survey of code systems will be given.

SECTION III.

KINDS OF ALPHABETS.

	Paragraph.
Definitions.....	13
Normal alphabet.....	14
Cipher alphabets.....	15
Standard alphabets.....	15a
Mixed alphabets.....	15b
Reciprocal alphabets.....	15c
Enciphering and deciphering alphabets.....	15d

13. Definitions.—Based upon an understanding of the foregoing principles, a few definitions will now be given.

An *alphabet* is an arbitrarily arranged sequence of elementary sounds to which an arbitrarily arranged sequence of symbols has been applied. Every alphabet consists, therefore, of two parts or *components*: (1) A sequence of sounds and (2) a sequence of symbols.

14. Normal alphabet.—The *normal alphabet* for any language is one in which the two components of the alphabet are the conventional sequences. The normal alphabets of the different languages vary somewhat as regards the number of elementary sounds and their arrangement in the normal sequence; our alphabet has 26, the French 25, the Italian 21, the Arabic 28, etc.

15. Cipher alphabets.—A *cipher alphabet*, or, as it is sometimes called, a *substitution alphabet*, is one in which the elementary or alphabetic sounds are represented by characters other than those representing them in the normal alphabet. These characters may be letters, figures, signs, symbols, or combinations of them. Except in the most amateurish cryptograms, symbol or sign ciphers are never encountered because such ciphers do not lend themselves to telegraphic transmission. The great majority of ciphers are composed

of letters, though number ciphers are frequently encountered. It will be convenient to designate that component of a cipher alphabet which consists of the sequence of sounds the *plain component*, and the component which consists of the sequence of symbols, the *cipher component*. In writing a cipher alphabet, if the plain component is omitted, the latter is understood to be the normal sequence. It will be convenient to indicate letters of the plain component, or the plain text, by suffixing the subletter "p" to them; letters of the cipher component, or the cipher text, by suffixing the subletter "c" to them. Thus, if we write $A_p = Q_c$, it means that plain-text letter A is represented by cipher letter Q, or vice versa, Q of the cipher represents A of the plain text. This will avoid all ambiguity in reference.

15a. Standard alphabets.—As regards the arrangement of the letters of the cipher component, cipher alphabets are of two kinds: (1) The sequence of letters in the cipher component may be the same as the normal sequence except that (a) the whole sequence has been shifted 1 to 26 letters, or (b) the sequence proceeds in the reversed direction from the normal, and may or may not be shifted; or (2) the sequence of letters may be altered from the normal. Alphabets of the first type are here designated as *standard cipher alphabets*; the second type, *mixed cipher alphabets*. In the former type, if the sequence proceeds in the same direction as the normal (in English, from left to right) we have a *direct standard alphabet*; if the sequence proceeds in the opposite direction from the normal we have a *reversed standard alphabet*.

Since our alphabet contains 26 elementary sounds, it is possible to apply the sequence of symbols to the sequence of sounds at 26 different points of coincidence, each yielding a different set of values for the sounds. If both sequences proceed in the same direction, then one of these points of coincidence corresponds, of course, to the normal, and the result is the normal alphabet, but the other 25 points of coincidence will all yield different direct standard alphabets. The following is one of the series of such alphabets:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—V W X Y Z A B C D E F G H I J K L M N O P Q R S T U

If the two sequences proceed in opposite directions, then we can apply them to each other at 26 different points of coincidence, each yielding different reversed standard alphabets. The following is one of the series of such alphabets:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—I H G F E D C B A Z Y X W V U T S R Q P O N M L K J

15b. Mixed alphabets.—Mixed alphabets may be of two kinds: (1) *Systematically mixed*, or (2) *random mixed*. The mixing of the

letters in the former kind is done according to some system which permits of reconstructing the alphabet at will. The purpose of this is to obviate the necessity of carrying written alphabets on the person and thus reduces the chances of loss or capture. The number of systems of mixing the letters is naturally very great, and we shall later illustrate a few of the most frequently encountered methods.

In a random-mixed alphabet the mixing is done absolutely haphazard, by drawing the letters out of a hat, or some such method whereby the laws of chance can operate to produce a thorough disarrangement. Such alphabets are safer from a cryptographic point of view because they afford no clues with regard to the probable positions of any letters, given the positions of a few of them, as is the case in systematically mixed alphabets. Their chief disadvantage is that they must be reduced to writing, since they can not be reconstructed at will from an easily remembered word or phrase. The following is an example of a random-mixed alphabet:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—R H E Q C W M A V N S G L B U Y F K D Z O X T I P J

15c. Reciprocal alphabets.—A *reciprocal alphabet* is one in which all the values in the alphabet are reciprocal in pairs. For example, $A_p = Q_o$ and $Q_p = A_o$ constitute a pair of reciprocal values. When this relation exists throughout the cipher alphabet the latter is termed a reciprocal alphabet. They are invariably produced when the two components are similar sequences but proceed in opposite directions. The example of a reversed standard alphabet above is also an example of a reciprocal alphabet. Often two or three pairs of values are found to be reciprocals in an otherwise nonreciprocal alphabet: in such cases the reciprocals are merely due to chance. Complete reciprocal alphabets can, however, be produced by arbitrary assignment of reciprocal values, but to do so serves no useful purpose. A knowledge or even a suspicion that the cipher alphabets used in a cryptogram that is being subjected to analysis are reciprocal alphabets is of great assistance in solution.

15d. Enciphering and deciphering alphabets.—An *enciphering alphabet* is one in which the sequence of symbols in the plain component coincides with the normal alphabet, for convenience in enciphering. A *deciphering alphabet* is the converse of its equivalent enciphering alphabet, that is, the sequence of symbols in the cipher component coincides with the normal alphabet, for convenience in deciphering. All the cipher alphabets shown above are arranged as enciphering alphabets. The following is the deciphering alphabet of the random-mixed alphabet shown above:

Cipher—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain—H N E S C Q L B X Z R M G J U Y D A K W O I F V P T

SECTION IV.

MONOALPHABET AND POLYALPHABET SUBSTITUTION CIPHERS.

	Paragraph.
Definition of terms.....	16
How to distinguish the two kinds of substitution ciphers.....	17

16. Definition of terms.--The simplest type of substitution cipher, called a *monoalphabet substitution cipher*, is the one in which a single cipher alphabet is employed throughout the whole message; that is, a given plain-text letter throughout the message is invariably represented by the same cipher letter, and conversely, a given cipher letter represents one and only one plain-text letter throughout the message. A *polyalphabet substitution cipher* is one in which a plurality of cipher alphabets is employed within the same message; that is, a given plain-text letter may be represented by several or many different cipher letters according to a definite rule governing the change, and conversely a given cipher letter may represent several or many different plain-text letters.

17. How to distinguish the two kinds of substitution ciphers.--A cryptogram which has been classed as a substitution cipher must then be examined to determine whether it is a monoalphabet or a polyalphabet substitution cipher. Usually this determination is easy to make, and is based upon the appearance of the graphic frequency table. The graphic table for a piece of normal plain text of fair length shows crests and troughs for two reasons: First, the sounds which the letters represent are used with greatly varying frequencies, and, second, each sound is invariably represented by the same letter. It follows, therefore, that since in a monoalphabet substitution cipher each different plain-text letter (=sound) is represented by one and only one cipher letter (=symbol), the graphic frequency table of such a cipher must also exhibit the irregular crest and trough appearance of the normal frequency table, but either the absolute positions or the relative sequence or order of the crests and troughs will be different from the normal. The marked irregularity in such a graphic table is in itself at once an indication that each symbol or cipher letter always represents the same plain-text letter. Hence the general rule: *A marked crest and trough appearance in the graphic frequency table indicates that a single alphabet is involved and is the test of a monoalphabet substitution cipher.* If, on the other hand, in a cryptogram each cipher letter represents several different plain-text letters, some of which are of high frequency, others of low frequency, it is clear that the graphic frequency table of the cipher will no longer exhibit the crest and trough appearance of the normal graphic table. For example, if in a cryptogram of a polyalphabetic nature K_c represents E_p in one position, G_p in another, J_p in another, and if in the same cryptogram R_c represents A_p in

one position, D_p in another, and B_p in another, the frequency of K_c and R_c will be approximately equal. The same would be true with respect to all the rest of the cipher letters in the message, with the result that the various points of the graphic table will be reduced to a more or less common level, and will give a "flattened" appearance to the table, because of this indiscriminate grouping of high and low frequency letters.

We are now ready for the analysis of a typical monoalphabet, or, as it is often called, single alphabet cipher.

SECTION V.

SOLUTION OF MONOALPHABET SUBSTITUTION CIPHERS USING STANDARD ALPHABETS.

Direct alphabet.....	Paragraphs.....	18
Reversed alphabet.....	19

18. Direct alphabet. (a) *Solution by frequency.* Let us first consider the case of standard monoalphabet ciphers. Standard alphabets are, as has been said, of two kinds, direct and reversed. The analysis of these two types of simple ciphers follows almost directly from a consideration of the nature of the cipher alphabets by means of which they are written. Since the sequence of letters in these cipher alphabets is the same as the normal sequence, the correct determination of the plain-text value of a single cipher letter in the cryptogram, whether the alphabet be direct or reversed, will result in the determination of the plain-text values of all the rest of the cipher letters at one stroke. The easiest point of attack is to assume that the cipher letter of highest frequency in the graphic table for the cryptogram represents E_p , and proceeding from this point, the values of the remaining letters are assigned on the basis of a direct or a reversed normal alphabet sequence. If the frequencies of the various cipher letters correspond in a general way with the normally to be expected frequencies of the plain-text letters they represent upon the assumption chosen, then the assigned values may be presumed to be correct. Substitution is made in the cryptogram, and if the analysis has been correct, solution results. If the original starting point—that is, the assumed value of E_p —is not correct, or if the direction of "reading" the successive points in the graphic table is not correct, then the frequencies of the various cipher letters will not correspond even approximately with the normally to be expected frequencies of the plain-text letters, and a new starting point is chosen, or the reversed direction of "reading" is applied. This process of applying the frequency relations of the normal graphic frequency table to a graphic frequency table of a cryptogram, in order to arrive at a solution, is spoken of as *fitting the graphic table to the normal*.

Let us apply these principles to the following cryptogram:

MESSAGE.¹

GJBMJ IUSJX YKBKT ZNJOB NUYZO RKLUX IKKYZ
 OSGZK JGYUT KKKMO SKTZO TLGTZ XEGTJ ZCUVR
 GZUUT YIGBG RKESU BOTMY UAZNU TMKZZ EYHAX
 MXUGJ NKGJU LIURA STTKG XOTMX UGJPA TIZOU
 TLOBK KOMNZ FKKUK GYZUL VOZFK XYINU URLOX
 KJAVU THEUA XVGZX URYNG BKJKY ZKUEK JHXOJ
 MKYUB KXSGX YNIXK KQLXU SMKKK TSUAT ZZUGV
 UOTZG YLGXT UXZNG YZNBH XOJMK JAKCK YZULV
 OZFKX YINUU RCORR JKLKT JNORR LOBKX OMNZY
 ODUTK SORKT UXZNU LMKKK TSUAT ZOLLU XIKJZ
 UXKZO XKCOR RJKYZ XUEHX OJMKY UAZNU LMKKK
 TSUAT ZGTJJ KRGEK KJYGZ SGXYN IXXXQ RUUSO
 YIGVZ

From the presence of so many low-frequency letters such as J, B, X, K, and Z, we at once suspect that this is a substitution cipher. But to make certain let us make a count of the letters by tabulating them in a graphic frequency table, shown in Figure 4.

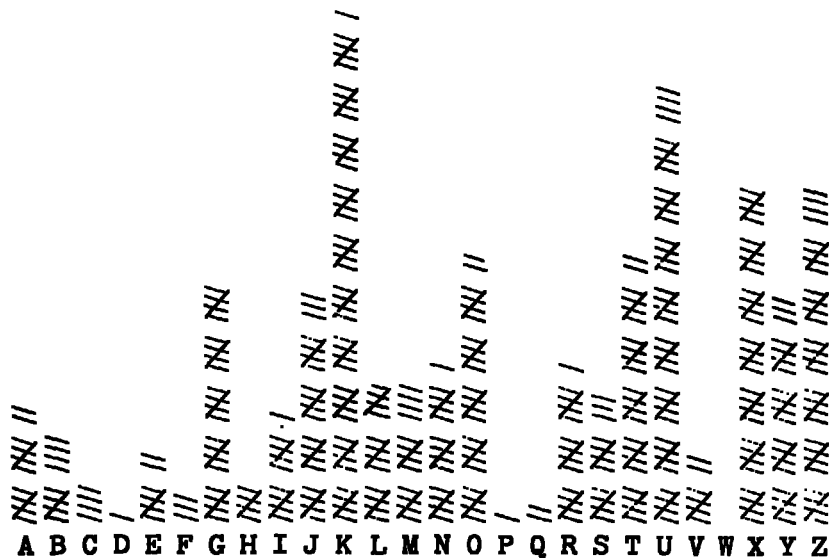


FIGURE 4.

Total number of letters = 425; vowels AEIOUY = 123 = 29 per cent; consonants HNRST = 78 = 18 per cent; consonants JKQXZ = 146 = 34 per cent.

¹ Cipher messages are usually sent in groups of five letters for two reasons: (1) It minimizes errors in telegraphic transmission, since an operator knows he must receive five letters, no more and no less, in each group; (2) it breaks up word lengths so that the enemy cryptanalysts gain no clues as to word formations, information which would, of course, be of great assistance to them in their attempts for solution.

Here we see that the vowels are 11 per cent below normal, the high-frequency consonants HNRST are 16 per cent below normal, and the low-frequency consonants JKQXZ are 32 per cent above normal. The cryptogram is therefore beyond all doubt a substitution cipher.

The next step is to determine whether it is a monoalphabet or a polyalphabet cipher. Referring to the test given on page 12, the decided irregularity of the graphic table, which shows marked crests and troughs, indicates that we are dealing with a monoalphabet substitution cipher. K_o being the greatest in frequency, we assume it to be E_p and proceed at once to fit the graphic table to the normal upon the basis of a direct standard alphabet. For example, if $K_o = E_p$, then, on the basis of a direct standard alphabet, $L_o = F_p$. Now F is a letter of medium or low frequency in the normal frequency table, and if our assumption, viz, that $K_o = E_p$, is correct, then L_o , which would represent F_p , should be of medium or low frequency; reference to Figure 4 shows this to be the case. On the other hand, O_o , which would represent I_p , should be a letter of high frequency, and we note that it is. The student is urged to go through the rest of the letters and their frequencies in the same way, in order to familiarize himself with this process, noting in particular how the crests fall at cipher letters O, T, U, X, Y, Z, and G (=I, N, O, R, S, T, and A, respectively), and the troughs fall at cipher letters P, Q, V, W, B, C, D, E, and F (=J, K, P, Q, V, W, X, Y, and Z, respectively). We find that upon the basis selected, namely, that the cryptogram is a monoalphabet, direct standard alphabet cipher, with $E_p = K_o$, an excellent fit of the graphic table to the normal can be made. The cipher alphabet is therefore as follows, arranged as an enciphering alphabet:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—G H I J K L M N O P Q R S T U V W X Y Z A B C D E F

Applying these values to the first few groups of the cryptogram we have the following:

Cipher—G J B M J I U S J X Y K B K T Z N J O B etc.
Plain—A D V G D C O M D R S E V E N T H D I V etc.

The deciphered message¹ is as follows:

ADV GD COMDR, 7TH DIV.

Hostile force estimated as one regiment infantry and two platoons cavalry moving south on GETTYSBURG ROAD. Head of column nearing road junction 580, east of PITZER SCHOOL fired upon by our patrols. Have destroyed bridges over MARSH CREEK from GREENMOUNT to a point as far north as the bridge due west of PITZER SCHOOL. Will defend hill 586 one mile north of GREENMOUNT.

If forced to retire will destroy bridge south of GREENMOUNT and delay Reds at MARSH CREEK.

LOOMIS,
Capt.

¹ The numbers which appear in the deciphered message were spelled out in the enciphered message. Sometimes the letters A to J are used for the digits, such cases being set off by a Q or an X, but this practice is not to be recommended on account of the possibility of errors.

(b) *Solution by completing the plain component.*—The foregoing method of analysis, involving as it does the process of fitting the graphic frequency table to the normal, may be called a "Solution by Frequency Table." There is, however, another method, much more rapid and really mechanical, which does not necessitate the compilation of a frequency table.

The enciphering alphabet shown directly above represents a case wherein the sequence of letters of both components of the cipher alphabet is the normal alphabet sequence but the cipher component has merely been shifted backward six letters, or forward twenty letters. If, therefore, we should take two normal sequences, regard them as the two components of the cipher alphabet, slide them against each other various numbers of letters, and apply the values given by each setting of the two sequences directly to the cryptogram, it is obvious that one of the points of application must yield the plain text. Thus, if we set the sliding components in this position:

Plain
ABCDEF GHIJK LMNOP QRSTUV WXYZ ABCDEF GHIJK LMNOP QRSTUV WXYZ
ABCDEF GHIJK LMNOP QRSTUV WXYZ
Cipher

and then apply the values given by this setting to the first three groups of the cryptogram, we have the following:

Cipher—GJBMJ	IUSJX	YKBKT
"Plain text"—HKCNK	JVTKY	ZLCLU

This does not yield intelligible text. We therefore shift the components one more letter apart and try again. Thus:

Plain
ABCDEF GHIJK LMNOP QRSTUV WXYZ ABCDEF GHIJK LMNOP QRSTUV WXYZ
ABCDEF GHIJK LMNOP QRSTUV WXYZ
Cipher

Cipher—GJBMJ	IUSJX	YKBKT
"Plain text"—ILDOL	KWULZ	AMDMV

This also does not yield intelligible text. But let us examine the results of the two trials.

Cipher—GJBMJ	IUSJX	YKBKT
Results of 1st trial—HKCNK	JVTKY	ZLCLU
Results of 2d trial—ILDOL	KWULZ	AMDMV

It is apparent that the net result of our experiments was merely to continue the normal alphabet sequence begun by each letter in

the several columns. It is obvious that if we complete the normal alphabet sequence in each column it will be exactly the same as continuing the successive trials of shifting the sliding components and applying the values to the cryptogram. Let us therefore complete the columns. We have what is shown in Figure 5. An examination of the successive horizontal lines, called *generatrices* (singular, *generatrix*), discloses one and only one line of plain text: ADVGDCOMDRSEVEN. Each column in Figure 5 is nothing but a series of direct alphabet sequences. Instead of laboriously writing down the several columns, we need merely to prepare a set of cardboard strips each bearing the normal alphabet repeated (to get coincidence at any setting), "set" the letters of our cryptogram upon one horizontal line, and then examine the successive generatrices for intelligible text. If the cryptogram is really one of this simple type, one of the generatrices will yield intelligible text, which will be the plain text of the message. This method of analysis is called a "Solution by Completing the Plain Component," and is one of the most valuable methods in cryptanalysis. It is recommended that the student prepare a set of 25 alphabets, each repeated so that every strip will contain 52 letters, and mount them upon strips of wood or some other material convenient to handle. Such a set of sliding alphabets will be found exceedingly valuable in all work of this kind.

19. Reversed alphabet. (a) *Solution by frequency table.*—The analysis of a cryptogram involving a single reversed standard alphabet is also very simple and yields to either the frequency table method of solution or the completion of plain component method. In fitting the graphic table to the normal in the case of a reversed alphabet the direction of "reading" the various points along the table is the reverse of that applied in the case of a direct alphabet. The following is an example:

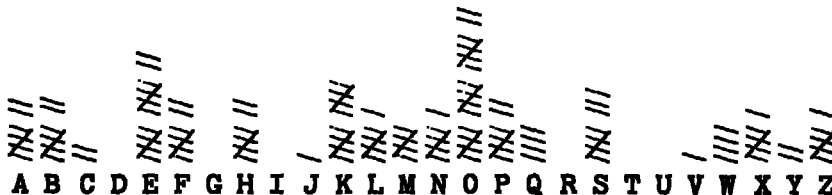
MESSAGE.

QMNKB AZQSX LSXOJ YAZEO SQLOP OSAZO BFOPM OENWE EPASH
 EPMNK XONKX OZWEV NKXOF KFONK XOBES PCLEA ZKHOA CYSB
 EFLEH PKFMM BXFKZ OLKHH WEEPA WLOOH OEQEH

GJBMJIUSJXYKBKT
 HKCNKJVTKYZLCLU
 ILDOLKWULZAMDMV
 JMEPMLXVMABNENW
 KNFQNMWYWNBCOFOX
 LOGRONZXOCDPGPY
 MPHSP OAYPDEQH QZ
 NQITQPBZQEF RIRA
 ORJURQCARFGSJSB
 PSKVS RDBSGHTKTC
 QTLWTSECTHIULUD
 RUMXUTFDUIJVMVE
 SVNYVUGEVJKWNWF
 TWOZV VHFVKLXOXG
 UXPAXWIGXLMYPYH
 VYQBYXJHYMNZQZI
 WZRCZYKIZNOARAJ
 XASDAZLJAOPBSBK
 YBTEBAMKBPQCTCL
 ZCUFCBNLCQRDUDM
 ADVGDCOMDRSEVEN
 BEWHEDPNESTFWFO
 CFXIFEQOFTUGXGP
 DGYJGFRPGUVVHYHQ
 EHZZKHGSQHVWIZIR
 FIALIHTRIWXJAJS

FIG. 5.

The graphic frequency table is as follows:



O_c is assumed to be E_p on account of its frequency. If we assume a direct alphabet then $P_c = F_p$, $Q_c = G_p$, etc. But the fit is very poor, for $Y_c = O_p$, and the frequency of Y_c is too low for O_p ; $Z_c = P_p$, and $A_c = Q_p$, which are both bad, being too high in frequency for these letters; and so on. But if we assume a reversed alphabet, still keeping $O_c = E_p$, an excellent fit is obtained. The cipher alphabet is as follows:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 Cipher—S R Q P O N M L K J I H G F E D C B A Z Y X W V U T

which when applied to the cryptogram yields the following:

CG, 1st Cav.

Have just reached eastern edge of woods along 552-595 road. Hostile squadron holding GRANITE HILL woods.

WHEELER,
 Col.

(b) *Solution by completing the plain component.*—The method of solution by completing the plain component may also be applied to a reversed standard alphabet cipher. The basic principles are the same as in the case of a direct standard alphabet cipher. We may experiment with two sliding components as before, only in this case one is the direct standard, the other the reversed standard. Let us set the two components against each other A to A, as shown below, and then attempt the decipherment of the first three groups:

Plain
 ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNPOQRSTUVWXYZ
 CIPHER
 ZYXWVUTSRQPONMLKJIHGFEDCBA

Cipher—QMNKB AZQSX LSXOJ
 "Plain text"—KONQZ ABKID PIDMR

This does not yield intelligible text. We therefore shift the reversed component one space forward and try again. Thus:

Plain
 ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNPOQRSTUVWXYZ
 CIPHER
 ZYXWVUTSRQPONMLKJIHGFEDCBA

Cipher—QMNKB AZQSX LSXOJ
 "Plain text"—LPORA BCLJE QJENS

Putting together the results of the two trials we have the following:

Cipher—QMNKBAZQSXLSXOJ
 Result of 1st trial—KONQZABKIDPIDMR
 Result of 2d trial—LPORABCLJEQJENS

It is seen that the letters of the *second* trial are merely the continuants of the normal alphabet sequences started by the letters of the first trial. If we complete these sequences in the several columns it will be the same as making the successive trials of shifting the components and trying for plain text in the cryptogram; one of the generatrices thus formed must therefore yield intelligible text. Let us see if it does. Figure 6 shows the various generatrices, and the plain text generatrix is underlined. The only difference in procedure between this case and the preceding (where the cipher alphabet was a direct standard) is that we set a reversed standard alphabet sequence QMNKBAZQSXLSXOJ against the normal and found the *plain component equivalents for the cipher letters before completing the plain component sequence started by each cipher letter.*

We shall find that in every case in which the cipher alphabet sliding components used in enciphering a message are known, this process of completing the plain component sequence can be applied to solve a message written by one or more unknown settings of the known components. It is only necessary to convert the cipher letters into their plain component equivalents before applying the completion process. In both of the cases above, the plain component was the normal alphabet, but in some subsequent cases we shall see that the plain components may even be mixed sequences. The completion sequence, however, must always be the plain component sequence in every case. It is immaterial at what points the plain and the cipher components are applied to each other in order to convert the cipher letters into their plain component equivalents.

FIGURE 6.

For example, in the case of the direct standard alphabet cipher on page 16, we arbitrarily set the cipher component sequence against the plain component sequence so that

$A_p = Z_c$. We might have set them at any one of the other 25 possible points of coincidence without affecting the final result, viz, the production of one plain text generatrix. Likewise, in the case of the reversed standard alphabet cipher above, we set the cipher component sequence against the plain component sequence so that $A_p = A_c$, though we might have set them at any one of the other 25 possible points of coincidence without affecting the final result.

SECTION VI.

SOLUTION OF MONOALPHABET SUBSTITUTION CIPHERS USING MIXED ALPHABETS.

	Paragraph
Reasons for greater difficulty in solving mixed alphabet ciphers.....	20
Digraphs and trigraphs.....	21
Trigraphic frequency table.....	22
Determining the repetition of digraphs and trigraphs in the message.....	23
Distinguishing the vowels from the consonants.....	24
Trial of substitution of deduced values.....	25
Finishing the solution—reconstructing the cipher alphabet.....	26
Solution of other messages using the same sliding alphabet.....	27

20. Reasons for greater difficulty in solving mixed alphabet ciphers.—We have seen thus far that in the case of standard alphabet ciphers the correct determination of the value of but one cipher letter results in the solution of the whole message. The reason for this is, of course, that the sequences of letters in both components of the cipher alphabet correspond to the sequences of letters in the normal alphabet and are therefore what may be termed *known sequences*. In a mixed alphabet, however, the sequence of letters in the cipher component of the cipher alphabet is different from the normal sequence, and is therefore an unknown sequence. It is necessary therefore to solve the values *individually*, since the value of one cipher letter gives no direct clues to the values of any of the other cipher letters. The solution of such a cipher involves considerably more analysis and experiment, therefore, than that of the ciphers we have heretofore been considering. Let us now consider the analysis of a typical example.

MESSAGE.

r	EMHTZ	LVDFG	SDRPS	FSDZF	IOGHL	PZFGZ	DYSPF
	HBZDS	GVHTF	UPLVD	FGYVJ	VFVHT	GADZZ	AITYD
A	ZYFZJ	ZTGPT	VTZBD	VFHTZ	DFXSB	GIDZY	VTXOI
	YVTEF	VMGZZ	THLLV	XZDFM	HTZAI	TYDZY	BDVFH
	TZDFK	ZDZZJ	SXISG	ZYGAV	FSLGZ	DTHHT	CDZRS
	VTYZD	OZFFH	TZAIT	YDZYG	AVDGZ	ZTKHI	TYZYS
	DZGHU	ZFZTG	UPGDI	XWGHX	ASRUZ	DFUID	EGHTV
	EAGML	OSTSE	STRSN	HD			

This message is obviously a substitution cipher. A brief examination discloses several repetitions, which have been underscored, a procedure which is usually a preliminary step to the analysis of any cryptogram. All attempts to solve this cipher by the method of completing the plain component sequence, on the assumption that either a direct or a reversed standard cipher alphabet is involved having failed, a graphic frequency table is made, and is shown in Figure 7.

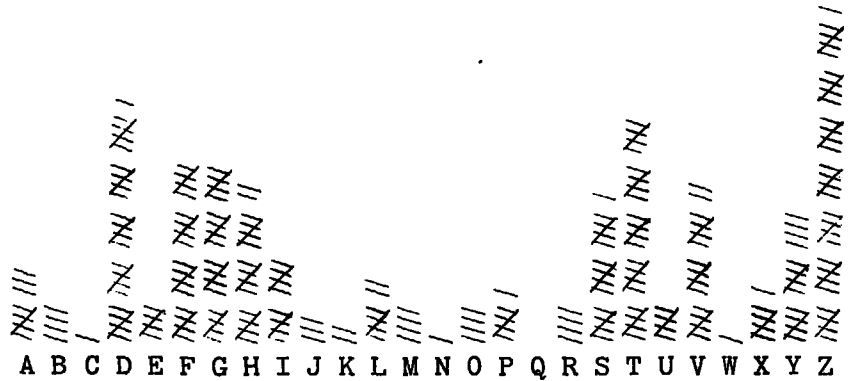


FIGURE 7.

The fact that the graphic table shows marked crests and troughs and has proved not to be a standard alphabet frequency distribution makes it practically certain that it involves a single mixed alphabet. We could, of course, attempt to solve it on the basis of individual frequencies of the cipher letters as compared with the normally to be expected frequencies. For example, we may assume that Z_o , the letter of greatest frequency, is E_p ; T_o , the letter of next greatest frequency, is O_p (assuming telegraphic text for the message), and so on. Of course if the message were long enough this method would yield the solution without difficulty; but the message is relatively short and we would encounter considerable difficulties due to the fact that these small frequencies probably differ to a greater or lesser degree from the normally to be expected frequencies.

21. Digraphs and trigraphs.—We may, however, bring to our aid certain other data with respect to normal frequencies. Just as the individual letters of plain text have more or less characteristic frequencies so we find that pairs of letters, called *digraphs*, and sets of three letters, called *trigraphs*, have characteristic frequencies. Hitt gives a table, based upon a count of 20,000 letters of military or semimilitary text, in which the normal occurrences of digraphs are reduced to a basis of 2,000 letters. Practically every combination can occur, but certain of them so rare that when the table is reduced to a basis of 2,000 letters, such cases would have to be assigned fractional values. They are therefore omitted from the table and may

be neglected. This table has been used in compiling the list of the first 30 most frequently recurring digraphs shown in Table 1. Below the list of digraphs is given the list of trigraphs to be expected in a count of 10,000 letters of English text and the frequency of single letters as initial and final letters of words, all as given by Hitt.

TABLE 1.

Digraphs to be expected in 2,000 letters of English literary text. Based on a count of 20,000 letters.

[Taken from Hitt's Manual.]

FIRST LETTER.

SECOND LETTER.	FIRST LETTER.																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1	7	10	22	3	2	26	4	2	2	7	8	11	2	9	13	12	9	2	4	1	12
B	5	...	1	2	1	...	1	1	1	2	2	1	3	...	1	3	...	4	1	12	...
C	6	...	1	1	14	2	...	11	11	3	...	2	3	1	1	...	1	...	1	...	1
D	6	...	12	30	1	2	...	4	...	30	1	...	4	1	1	1	...	1	...	1	...	3
E	11	14	16	12	2	6	33	10	2	6	18	14	12	1	7	36	11	12	2	16	5	...	1	1	1	1
F	3	...	2	8	2	1	...	2	...	2	1	3	25	3	1	1	...	1	...	1	...	1
G	4	...	1	3	2	11	2	...	3	1
H	1	...	11	2	4	1	4	...	1	...	1	2	1	1	2	10	50	...	3	3	...	2
I	2	1	4	12	6	5	1	12	1	...	5	9	8	12	1	3	12	13	22	2	3	6	...	1	1	1
J	1
K	1	1	...	2	2	1	2	1	...	1
L	14	6	2	1	6	1	1	6	...	9	...	3	6	3	3	2	3	5
M	7	...	3	13	2	...	2	3	...	4	1	10	...	4	1	1	2
N	38	...	3	25	...	2	1	31	...	3	...	2	2	39	...	4	3	11	...	2
O	1	1	12	4	8	8	3	12	18	2	...	4	7	8	3	7	13	15	22	...	2	6	1	5
P	2	...	1	8	1	...	2	4	2	3	2	1	8	1	4	...	3	1	...	3	1
Q	2	1	1	...	1	1
R	16	1	3	3	40	3	6	2	6	...	1	2	1	25	8	2	2	8	11	2
S	16	1	...	3	25	1	2	...	17	...	1	2	1	12	7	2	9	11	6	11	...	1	...	6
T	25	1	3	12	13	5	2	3	20	...	2	1	24	8	2	16	20	11	6	...	2	2	7
U	1	2	1	6	1	3	2	2	...	3	1	...	17	1	5	3	5	5	1
V	3	1	...	5	5	3	2	...	3	2	...	2	5	...	5	...	1
W	1	...	2	8	...	1	1	1	1	2	4	2	3	3
X	1	...	4	2	1	...	1
Y	3	2	...	2	4	...	1	1	8	1	2	...	1	3	1	7
Z	1	1	1

Most frequent digraphs.

- | | | |
|-------|-------|-------|
| TH—50 | AT—25 | ST—20 |
| ER—40 | EN—25 | IO—18 |
| ON—39 | ES—25 | LE—18 |
| AN—38 | OF—25 | IS—17 |
| RE—36 | OR—25 | OU—17 |
| HE—33 | NT—24 | AR—16 |
| IN—31 | EA—22 | AS—16 |
| ED—30 | TI—22 | DE—16 |
| ND—30 | TO—22 | RT—16 |
| HA—26 | IT—20 | VE—16 |

Most frequent trigraphs.

THE—89	TIO—33	EDT—27
AND—54	FOR—33	TIS—25
THA—47	NDE—31	OFT—23
ENT—39	HAS—28	STH—21
ION—36	NCE—27	MEN—20

Frequency of initial and final letters.

Letters—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Initial —9 6 6 5 2 4 2 3 3 1 1 2 4 2 10 2 - 4 5 17 2 - 7 - 3 -
Final —1 - - 10 17 6 4 2 - - 1 6 1 9 4 1 - 8 9 11 1 - 1 - 8 -

Relative frequencies of the vowels.

A 19.5% E 32.0% I 16.7% O 20.2% U 8.0% Y 3.6%

Average number of vowels per 20 letters, 8.

22. Trigraphic frequency table.—If we are to apply these data to our problem we must compile a TRIGRAPHIC FREQUENCY TABLE for the cryptogram, to show what digraphs and trigraphs occur in it. This table is shown in Figure 8. It is convenient to use cross-section paper for this work. The letter which precedes a given letter is termed its *prefix*, and the one which succeeds the given letter, its *suffix*. In the trigraphic table, Figure 8, the upper line of each pair of lines opposite the letters of the normal alphabet gives the prefixes to the cipher letters, the lower line of each pair gives the suffixes; the prefix and suffix to each letter thus occupy corresponding positions in each pair of lines. The first letter of our message to be tabulated is E. It has no prefix, and therefore a dash is placed in the space that would otherwise be occupied by a prefix; its suffix is M, and the letter M is written beneath the dash. The next letter to be tabulated is M; its prefix, E, and suffix, H, occupy the first positions opposite M in the table. The next letter to be tabulated is H; its prefix, M, and suffix, T, are written in the table opposite H, and so on until all the letters have thus been tabulated. The last letter of the message, D, has no suffix, and a dash indicates this fact in the table.

23. Determining the repetition of digraphs and trigraphs in the message.—We now proceed to find the digraphs and trigraphs which are repeated in the message. Consider the letter D, which shows the letter Z to occur nine times as a prefix to D, which means that the digraph ZD occurs nine times in the cryptogram. The letter F is also indicated as a suffix to D six times, which means that the digraph DF occurs six times in the cryptogram. We may determine the repeated digraphs by considering either the prefixes to the letters or the suffixes, the same final results being obtained in either case so long as consistency is maintained in listing the digraphs.

24

ELEMENTS OF CRYPTANALYSIS.

A GZZGZGXE
 DIIVIVSG
 B HZSY
 ZDGD
 C T
 D
 D VSSZZVAYBZIZYBZZZCZYVSGZIH
 FRZYSFZZVFZFZVFZTZOZGZIFE-
 E -TDVS
 MFGAS
 F DSZZPTDVYVDEDVDVZFZD
 GSIGHUGVZHXVMHKSFHZU
 G FOFSTTBMSYLYDZTPWEA
 SHZVYAPIZZAZAZHUDHMM
 H MGFVVFTMFTHFKGGGN
 TLBTTTLTTHTTTIUXTD
 I FAGOAXAHDU
 OTDYTSTTXD
 J VZZ
 VZS
 K FT
 ZH
 L ZPHLSM
 VPVLVGO
 M EVFG
 HGHL
 N S
 H
 O IXDL
 GIZS
 P RLSUGU
 SZFLTG
 Q
 R DZST
 PSUS

Condensed table of repetitions.

HTZ-5	HT-9	VT-4
ITY-4	ZD-9	ZZ-4
ZDF-4	DZ-9	FH-4
DZY-4	DF-6	GH-4
ZAI-3	TZ-6	IT-4
YDZ-3	ZY-6	VF-4
	TY-5	ZF-4
	GZ-5	ZT-4

FIGURE 8.—Trigraphic frequency table.

S	GPFYDXJIFRYAOTER DFDPGBXGLVDR TETN
T	HHHIZPVHV VZHIHDHVHIZIZHSS ZFGYGVZZXEHZY ZHCYZYKYGVSR
	FHGRF PZPZI
V	LGLYJFTDY YFLDASAT DHDJFHTFTT MXFFTDE
W	X G
X	FTVSIH SOZIWA
Y	DGTZZITZZTTZTZ SVDFVVDBGZDGZS
Z	TDPGBDZDFJTTDGZXTDTKDZGGDYOTDGZYDUFU LFFDDZAYJTB DYZTDA YDDZJYDRDFAYZTYGFTD

FIGURE 8.—Trigraphic frequency table (cont'd).

To determine the repeated trigraphs we must find those cases in which two or more prefixes to a given letter are identical at the same time that the suffixes to the same letter are identical. For example, the table shows that in four cases the prefix to the letter D is the letter Z at the same time that the suffix to this letter is the letter F. Hence, the trigraph ZDF occurs four times. The repeated trigraphs are all determined in this manner. The most frequently repeated digraphs and trigraphs are noted in a condensed table, so as to bring this important information prominently before the eye. Digraphs which occur less than four or five times and trigraphs which occur less than three or four times may be omitted from the condensed table, for they are relatively nonsignificant.

24. Distinguishing the vowels from the consonants.—Before proceeding to an analysis of these digraphs and trigraphs, let us make a list of the 10 cipher letters of greatest frequency and see if we can determine which of them represent vowels and which represent consonants.

Frequency—	36	26	25	20	20	17	17	16	14	10
Letter—	Z	D	T	F	G	H	V	S	Y	I

There can be hardly any doubt but that $Z_0 = E_p$, and we will assume this to be the case. If an examination is made of the list of most

frequently repeated digraphs given in connection with Table 1, it will be seen that a combination of two vowels does not appear among the first 15 digraphs given. All but two of these digraphs are in fact combinations of the high-frequency vowels with the high-frequency consonants. Of these 13 vowel and consonant digraphs, the vowel E enters into six pairs, the vowels O and A into three pairs each, the vowel I into only one pair. The combinations of vowel E with the high-frequency consonants are in fact predominant. If, therefore, we indicate in the list of the first 10 letters of greatest frequency in our problem the number of times Z_o enters into combination with each of these 9 other letters, we shall have some indication as to which of them probably represent vowels, which consonants. The figures above the letters indicate the number of times Z_o occurs as a prefix to the letters; the figures below the letters indicate the number of times Z_o occurs as a suffix to the letters.

Z_o as prefix—9	4	4	1	0	0	0	6	0
Letters—D	T	F	G	H	V	S	Y	I
Z_o as suffix—9	6	2	5	0	0	0	2	0

From these data we may conclude at once that D, T, F, G, and Y are in all probability consonants, whereas H, V, and S are in all probability vowels. The letter I_o may be a vowel too, but since the four vowels A, E, I, and O are much more frequent than U, and since we have classified the four letters Z_o , H_o , V_o , and S_o as vowels, we may leave I_o as unclassified for the present.

Let us now list all the combinations of Z, H, V, and S that occur in the message as well as the number of times they occur.

ZZ—4 HH—1 VH—2 SV—1

The doublet ZZ_o is EE_p , a fairly frequent occurrence in English. We have another doublet, HH_o , which is probably OO_p , for AA_p and II_p are hardly ever encountered in English. If $H_o = O_p$, then either V_o or S_o is I_p . Now VH_o occurs twice, and if $V_o = I_p$, then $VH_o = IO_p$, a very common digraph. Moreover, VH_o is both times followed by T_o , which has already been classified as a consonant and would be very good for N_p , giving the high-frequency trigraph ION_p as the equivalent of VHT_o . If we are correct thus far, then S_o , the last of the four letters classified as vowels, must be A_p . We now have the following assumed values:

$Z_o = E_p$ $T_o = N_p$ $H_o = O_p$ $V_o = I_p$ $S_o = A_p$

The letter D_o , second highest in frequency, is a consonant. The reversible digraphs ZD and DZ, each occurring 9 times, would be excellent for ER and RE. The only high-frequency letters for which

we have not yet assumed values are T_p and S_p . The letters F_o and G_o have been classified as consonants, and we may be fairly certain that they represent these two high-frequency consonants, but we can not distinguish between them from our data. For example, the digraphs in which F_o and G_o occur would lend equal weight to the assumptions that $F_o = T_p$ or $G_o = T_p$:

$$\begin{array}{l} 6\text{-DF} = \begin{cases} \text{RT} \\ \text{RS} \end{cases} \quad 4\text{-FH} = \begin{cases} \text{TO} \\ \text{SO} \end{cases} \quad 4\text{-GH} = \begin{cases} \text{SO} \\ \text{TO} \end{cases} \quad 5\text{-GZ} = \begin{cases} \text{TE} \\ \text{SE} \end{cases} \\ 4\text{-VF} = \begin{cases} \text{IT} \\ \text{IS} \end{cases} \quad 4\text{-ZF} = \begin{cases} \text{ER} \\ \text{ES} \end{cases} \quad 3\text{-TG} = \begin{cases} \text{NT} \\ \text{ST} \end{cases} \end{array}$$

We may, however, give the letters F_o and G_o alternate values, S_p and T_p . We now have the following values:

$$Z_o = E_p, T_o = N_p, H_o = O_p, V_o = I_p, S_o = A_p, D_o = R_p, F_o = \begin{cases} T_p \\ S_p \end{cases}, G_o = \begin{cases} S_p \\ T_p \end{cases}$$

25. Trial of substitution of deduced values.—We have not as yet substituted in the cryptogram a single one of the values deduced from our analysis of the trigraphic frequency table, and it may be that we have really gone too far with the process. The test of the correctness of our assumptions is after all only this: Do they yield "skeletons" of words in the cryptogram? Let us see. Here are the results of substituting the eight values deduced by our reasoning:

EMHTZ LVDFG SDRPS FSDZF IOGHL PZFGZ DYSPF HBZDS GVHTF
 ONE IRTS AR A TARET SO ETSE R A T O ERA SIONT
 ST S S T ST S T S

UPLVD FGYVJ VVHT GADZZ AITYD ZYFZJ ZTGPT VTZBD VFHTZ
 IR TS I ITION S REE N R E TE ENS N INE R ITONE
 ST S T S T S

DFXSB GIDZY VFXOI YVTEF VMGZZ THLLV XZDFM HTZAI TYDZY
 RT A S RE IN IN T I SEE NO I ERT ONE N RE
 S T S T S

BDVFH TZDFK ZDZZJ SXISG ZYGAV FSLGZ DTHHT CDZRS VTYZD
 RITO NERT EREE A AS E S I TA SE RNOON RE A IN ER
 S S T T S T

OZFFH TZAIT YDZYG AVDGZ ZTKHI TYZYS DZGHU ZFZTG UPGDI
 ETTO NE N RE S IRSE EN O N E A RESO ETENS SR
 SS T T T S T T

XWGHX ASRUZ DFUID EIGHTV EAGML OSTSE STRSN HD
 SO A E RT R SONI S ANA AN A OR
 T S T T

Several long words, nearly complete, stand out. Note the following underlined portions in the message:

HBZDSGVHTF	TVTZBDVFHTZDF	SLGZDTHHT
O ERASIONT	NINE RITONERT	A SERNOON
T S	S S	T

Here we have the words "OPERATIONS," "NINE PRISONERS," and "AFTERNOON." The value of G_o is clearly T_p ; that of F_o is S_p . It is clear that B_o equals P_p in both cases, and its frequency is excellent for P_p . Also L_o is good for F_p .

26. Finishing the solution — reconstructing the cipher alphabet.—Substitution of the new values obtained by completing the skeletons of words results in speedy reduction of the cryptogram. The final message is as follows:

EMHTZ LVDFG SDRPS FSDZF IOGHL PZFGZ DYSPF HBZDS GVHTF
 GXONE FIRST ARMYA SARES ULTOF YESTE RDAYS OPERA TIONS
 UPLVD FGYVJ VEVHT GADZZ AITYD ZYFZJ ZTGPT VTZBD VFHTZ
 BYFIR STDIV ISION THREE HUNDR EDSEV ENTYN INEPR ISONE
 DFXSB GIDZY VTXOI YVTEF VMGZZ THLLV XZDFM HTZAI TYDZY
 RSCAP TURED INCLU DINGS IXTEE NOFFI CERSX ONEHU NDRED
 BDVFH TZDFK ZDZZJ SXISG ZYGAV FSLGZ DTHHT CDZRS VTYZD
 PRISO NERSW EREEV ACUAT EDTHI SAFTE RNOON QREMA INDER
 OZFFH TZAIT YDZYG AVDGZ ZTKHI TYZYS DZGHU ZFZTG UPGDI
 LESSO NEHUN DREDT HIRTE ENWOU NDEDA RETOB ESENT BYTRU
 XWGHX ASRUZ DFUID EGHTV EAGML OSTSE STRSN HD
 CKTOC HAMBE RSBUR GTONI GHTXF LANAG ANMAJ OR

G-1, FIRST ARMY.

As a result of yesterday's operations by 1st Division, 379 prisoners captured, including 16 officers. 100 prisoners were evacuated this afternoon. Remainder less 113 wounded are to be sent by truck to Chambersburg to-night.

FLANAGAN,
 Major.

It will be useful to reconstruct the cipher alphabet employed by these correspondents, *arranging it as an enciphering alphabet.*¹ It is as follows:

Plain—	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—	S U X Y Z L E A V N W O R T H B C D F G I J K M P Q

¹ It is always best to arrange the cipher alphabet as an enciphering alphabet, because it will then be in the form in which the encipherer employed it, in which case, if the sequence of letters in the cipher component shows any system of construction or derivation, valuable clues to the analysis of similar cryptograms between the same correspondents will be gained. The deciphering alphabet will not show such clues.

We note that the cipher component is a *key-word alphabet*, based upon the word LEAVENWORTH. The cipher component is six letters in advance of A of the plain component, that is, L, the first letter of the key-word sequence, is set beneath F, the sixth letter of the normal sequence.

The example solved above is admittedly a more or less artificial illustration of the method of analysis, made so in order to demonstrate in a general way the usual procedure. It was easy to solve because the frequencies of the various cipher letters corresponded rather well with the normally to be expected frequencies. However, all cryptograms of the same monoalphabetical nature, using a cipher alphabet in which the cipher component is a mixed sequence, can be solved along the same general lines after more or less experimentation.

27. Solution of other messages using the same sliding alphabet.—Once the cipher alphabet has been reconstructed, subsequent messages enciphered by means of the same basic mixed alphabet may be solved very readily. We have seen that the preceding message was enciphered by sliding the cipher component LEAVN WORTHBCD . . . XYZ six letters to the right of A of the plain component. It is obvious, of course, that the cipher component may be set against the plain component at any one of 26 different points of coincidence, each yielding a different cipher alphabet. Since, however, the cipher component has been reconstructed by our analysis above, it has become a *known sequence*, and we are therefore enabled to apply the method of completing the plain component to any subsequent cipher messages enciphered by means of this basic cipher component set at any one of the 26 possible points of coincidence. An example will serve to make the process clear. Let us suppose that we have intercepted the following message passing between the same two stations as before:

MESSAGE.

IYEWK CERNW OFOSE LFOOH EAZBI FN NYO

We first convert the first two or three groups of cipher letters into their plain component equivalents by setting the cipher component beneath the normal alphabet, and then use our normal alphabet sliding strips to complete the normal alphabet sequence beneath each column. Thus:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher—	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z

Cipher—	I Y E W K C E R N W
Plain equivalents—	P Y B F R L B H E F
	Q Z C G S M C I F G
	R A D H T N D J G H
	S B E I U O E K H I
	T C F J V P F L I J
	U D G K W Q G M J K
	V E H L X R H N K L
	W F I M Y S I O L M
	X G J N Z T J P M N
	Y H K O A U K Q N O
	Z I L P B V L R O P
	A J M Q C W M S P Q
	B K N R D X N T Q R
	C L O S E Y O U R S
	D M P T F Z P V S T
	E N Q U G A Q W T U
	F O R V H B R X U V
	G P S W I C S Y V W
	H Q T X J D T Z W X
	I R U Y K E U A X Y
	J S V Z L F V B Y Z
	K T W A M G W C Z A
	L U X B N H X D A B
	M V Y C O I Y E B C
	N W Z D P J Z F C D
	O X A E Q K A G D E

Note the plain text generatrix CLOSE YOURS.

We may solve the rest of the message in exactly the same way, or we may set the cipher component beneath the normal alphabet so that $C_p = I_c$ (the value obtained for the first cipher letter as a result of our solution of the first two groups) and solve the rest of the message directly from the cipher alphabet itself. It is as follows:

Close your station at 2 P. M. CARR,
Lt.

FIGURE 9.

The cipher alphabet for this message is as follows:

Plain—	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher—	F G I J K M P Q S U X Y Z L E A V N W O R T H B C D

By merely shifting the cipher component, the whole series of values has been changed at one stroke from those in the long message above, but, nevertheless, solution of the very short message in a different key was obtained, as we have seen, very readily, *without any frequency table analysis*.

SECTION VII.

REMARKS ON THE SOLUTION OF MONOALPHABET CIPHERS.

	Paragraph.
Cryptograms using numbers or symbols.....	28
Examples and solution of historic ciphers.....	29

28. Cryptograms using numbers or symbols.—Monoalphabet ciphers, we have seen, are very easy to solve. In the case of single standard alphabet ciphers not even frequency tables are necessary; where the cipher alphabet is a mixed alphabet, a trigraphic frequency table is necessary to solution, but subsequent cryptograms written by means of the same cipher component, although set according to different key letters, may be solved without recourse to frequency tables.

Any monoalphabet substitution cipher written by means of a mixed alphabet, whether the cryptogram consists of letters, figures, symbols, or combinations of them, may be solved by the method illustrated above. In dealing with cryptograms composed of numbers or signs it is usually of advantage for practical purposes to substitute arbitrary letters for the numbers or symbols consistently throughout the message and proceed as usual.

29. Examples and solution of historic ciphers.—Two examples of historical interest will be cited in this connection as illustrations. During the campaign for the presidential election of 1876 many cipher messages were exchanged between the Tilden managers and their agents in several states where the voting was hotly contested. Two years later the New York Tribune exposed many irregularities in the campaign by publishing the decipherments of many of these messages. These decipherments were achieved by two investigators employed by the Tribune, and the plain text of the messages showed conclusively that illegal attempts and measures to carry the election for Tilden were made by his managers. Here is one of the messages:

JACKSONVILLE, *Nov. 16 (1876).*

GEO. F. RANEY, *Tallahassee:*

Ppyyemnshyyypimashnsyys sitepaae
 nshnsseusshnsmmpiyysnppyeaapieissyey
 hainssspeeiyshnynsssyepiaanyitnsshy
 yspyypinsyys site me ipimmeisseiyyeiss
 itelepyypeeiaaassimaayespnsyyian
 ssseissmppnspinsnpinsimimyyitemyys
 speyymmnsyys site spypeepppmaaayypit

L'Engle goes up to-morrow.

DANIEL.

Examination of the message discloses that only ten different letters are used. It is probable, therefore, that what we have here is a cipher in which the combinations of two letters represent single letters of the plain text. We therefore rewrite the message in pairs and substitute arbitrary letters for the pairs, as seen below:

PP YY EM NS HY YY PI MA SH NS YY SS etc.
A B C D E B F G H D B I etc.

A trigraphic frequency table is then made and analysis of the table along the lines illustrated above yields solution, as follows:

JACKSONVILLE, Nov. 16.

GEO. F. RANEY, Tallahassee:

Have Marble and Coyle telegraph for influential men from Delaware and Virginia. Indications of weakening here. Press advantage and watch Board. L'Engle goes up to-morrow.

DANIEL.

The other example, using numbers, is as follows:

JACKSONVILLE, Nov. 17.

S PASCO and E. M. L'ENGLE:

84	55	84	25	93	34	82	31
31	75	93	82	77	33	55	52
93	20	90	66	77	65	33	84
63	31	31	93	20	82	33	66
52	48	44	55	42	82	48	89
42	93	31	82	66	75	31	93

DANIEL.

There were, of course, several messages of like nature, and examination disclosed that only 26 different numbers in all were used. Solution of these ciphers followed very easily, the decipherment of the one given above being as follows:

JACKSONVILLE, Nov. 17.

S. PASCO and E. M. L'ENGLE:

Cocke will be ignored, Eagan called in. Authority reliable.

DANIEL.

The Tribune experts gave the following alphabets as the result of their decipherments:

AA=O	EP=C	MM=G	PI=R	SS=N
AI=U	IA=K	NN=J	PP=H	YE=F
EI=I	IM=S	NS=E	SH=L	YI=X
EM=V	IT=D	NY=M	SN=P	YY=A
EN=Y	MA=B	PE=T	SP=W	
20=D	34=W	52=U	75=B	89=Y
25=K	39=P	55=O	77=G	93=E
27=S	42=R	62=X	82=I	96=M
31=L	44=H	66=A	84=C	99=J
33=N	48=T	68=F	87=V	

They did not attempt to correlate these alphabets, or at least they say nothing about a possible relationship. The present author has, however, reconstructed the rectangle upon which these alphabets are based, and it is given herewith:

		<i>2nd letter or number</i>									
		H I S P A Y M E N T									
		1 2 3 4 5 6 7 8 9 0									
<i>1st letter or number</i>	H 1										
	I 2				K		S			D	
	S 3	L		N	W					P	
	P 4		R		H				T		
	A 5		U		O						
	Y 6	X				A		F			
	M 7				B		G				
	E 8	I		C			V		Y		
	N 9		E			M			J		
	T 0										

FIGURE 10.

It is amusing to note that the conspirators selected as their key a phrase quite in keeping with their attempted illegalities: HIS PAYMENT; for bribery seems to have played a considerable part in that campaign. The blank squares in the diagram probably contained proper names, numbers, etc.

Rectangles of the general nature of that shown above are commonly used. Sometimes the letters at the top and side of the rectangle are based upon key words, both of which are often the same, often different. The letters within the rectangle are often in the normal alphabet sequence; often they are mixed. Regardless of how the cipher alphabets are produced, whether by sliding sequences, or rectangles, or any other device, solution of monoalphabet substitution ciphers follows along the lines indicated, and it is always valuable to reconstruct the alphabets which were used, because of the clues they afford in the analysis of subsequent messages.

In the recent war the cases where monoalphabet ciphers were encountered in actual operations were exceedingly rare because of the simplicity of their solution. However, a few cases did occur, and one rather illuminating instance may be cited. In a rather important communication to Helfferich on August 5, 1918, General Kress von Kressenstein used a single mixed alphabet, and the intercepted radio message was solved at American G. H. Q. very speedily. A day later another message, but in a very much more difficult system of cipher, was intercepted and solved. It read when translated as follows:

G. H. Q. KRESS:

The cipher prepared by General von Kress was at once solved here. Its further use and employment is forbidden.

CHIEF SIGNAL OFFICER, Berlin.

SECTION VIII.

INTRODUCTORY REMARKS ON POLYALPHABET SUBSTITUTION
CIPHERS.

	Paragraph.
Object of using several alphabets in one message.....	30
Independent and interrelated cipher alphabets.....	31
Primary components and secondary alphabets—interrelated alphabets.....	32
Cipher squares or tables—Vigenère Table.....	33
Periodic and nonperiodic systems.....	34

30. Object of using several alphabets in one message.—In each of the cryptograms considered up to this point we have noted that only one cipher alphabet was employed throughout the message and that such ciphers are very easily solved because the frequencies of the cipher characters correspond to a greater or lesser degree with the frequencies of the letters of normal plain text. We come now to the more complicated types of substitution ciphers, in which several cipher alphabets are employed in the same message for the purpose of eliminating the frequency characteristics of the symbols employed in the cryptogram. The number of such systems is very great, and it will be possible to discuss only a few of the more common and simpler types of these polyalphabet ciphers. Some of the more complex types of polyalphabet ciphers require considerable experience and patience on the part of the cryptanalyst, for they entail a great amount of painstaking analysis and labor.

Before proceeding to a discussion of the methods of analysis of polyalphabet ciphers, it will be of advantage to point out the various types of cipher alphabets that are usually encountered in these cryptograms, how they are produced, and how the method of production may be made to yield clues to the analysis of the alphabets.

31. Independent and interrelated cipher alphabets.—A primary classification of cipher alphabets into two types may be made: (1) Independent or unrelated cipher alphabets, and (2) dependent or interrelated cipher alphabets.

(a) The first type may be disposed of in a few words. They are obtained by making up a number of mixed alphabets showing no relation to one another in any way, by drawing letters out of a hat, or by random assignments in enciphering alphabets. The solution of cryptograms written by means of such alphabets is rendered more difficult by reason of the absence of any definite relations between the equivalents of one cipher alphabet and those of any of the other cipher alphabets in the cryptogram. The analysis of such cases will be discussed later.

(b) The second type of cipher alphabets, viz, the one in which the various alphabets in a message are interrelated, calls for more detailed explanation. We have seen that in the case of monoalphabet ciphers the cipher alphabets are often produced by sliding two sequences of letters against one another, resulting in the production of a series of cipher alphabets. These sliding sequences were termed "components," and we saw that they could be normal or mixed se-

quences. It will be useful at this point to go further into detail with regard to the results of the sliding of such components.

32. Primary components and secondary alphabets—inter-related alphabets.—The two basic or fundamental sequences comprising a cipher alphabet have heretofore been designated as the plain and the cipher components. For our present purposes they will be termed the PRIMARY COMPONENTS, and the resultant cipher alphabets produced by sliding them against each other and arranging the equivalents in the form of enciphering or deciphering alphabets will be termed the SECONDARY ALPHABETS. Two components of 26 letters each will yield 26 secondary alphabets. The following types of primary components may exist:

a. The components are both normal sequences.

- (1) The sequences proceed in the same direction. The secondary alphabets are direct standard alphabets.
- (2) The sequences proceed in opposite directions. The secondary alphabets are reciprocal reversed standard alphabets.

b. The components are not both normal sequences.

- (1) One of the components, usually the plain component, is the normal sequence; the other, a mixed sequence. The secondary alphabets are mixed alphabets.
- (2) Both components are mixed sequences.
 - (a) Sequences are identical mixed sequences.
 1. Sequences proceed in the same direction. The secondary alphabets are mixed alphabets.
 2. Sequences proceed in opposite directions. The secondary alphabets are reciprocal mixed alphabets.
 - (b) Sequences are different mixed sequences. The secondary alphabets are mixed alphabets.

33. Cipher squares or tables—Vigenère Table.—Let us study the first case, viz, that in which the primary components are both normal sequences and proceed in the same direction, yielding direct standard alphabets. These secondary alphabets when tabulated in the form shown in Table 2 yield a table known in the literature of the subject under various names: "Vigenère Table," "Square Table," "Quadricular Table," "Pythagorean Table," etc. Such a table may be used in various ways, differing from each other only in minor details, but the most common one is to consider the top line as the plain-text line of letters, and the successive horizontal lines, the lines of cipher equivalents. When these lines are considered in conjunction with the top line, we have a series of cipher alphabets, each of which may be designated by the initial letter of the cipher line. Thus, the D, or 4th cipher alphabet, is the one in which $A_p = D_c$, $B_p = E_c$, and so on.

Therefore, when a series of such cipher alphabets is to be employed in a message, a key word or a key number can be agreed upon by the correspondents to determine the choice of the cipher alphabets, and the letters of the key word or the figures of the key number designate the particular line from which the cipher equivalents are to be taken in each case. This, of course, amounts to exactly the same thing as setting A of the plain component to those letters of the cipher component indicated by the key word when sliding sequences are used instead of the table.

Minor modifications of the Vigenère Table are encountered. If the top line is made a reversed normal sequence, or if the successive interior lines are made reversed normal sequences, then the secondary alphabets are all reversed standard alphabets, and are exactly the same as those produced by sliding the direct normal sequence against the reversed normal. Such a table is often referred to as the "Beaufort Table," named after an English Admiral who "invented" it, and thought he had discovered something new, but alphabets similar in their nature to the alphabets produced by the Beaufort Table were known and used over a thousand years before his time.

TABLE 2.

The Vigenère Square.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

The second type of sliding components listed above, viz, those which are not both normal sequences, calls for more detailed treatment and will be discussed later. Suffice it to say at this point that either sliding components or a square table may be used to produce the various cipher alphabets, and since these secondary alphabets are all interrelated, the existence of certain relationships in them is of great aid in the analysis of these alphabets. Let us proceed now to an illustration of how the various alphabets are used in a poly-alphabet cipher.

34. Periodic and nonperiodic systems.—Polyalphabet systems may be divided into two more or less distinct types: (1) Periodic systems and (2) nonperiodic systems. When the text of a cryptogram can be broken up into definite, regular groups, blocks, or cycles of letters which have undergone encipherment by identical portions of the key, the cryptogram is said to exhibit *cyclic phenomena*, and to be of the *periodic* type. If the text of a cryptogram can not thus be treated, and does not exhibit any cyclic phenomena, it is said to be of the *nonperiodic* type.

We shall first take up the ciphers of the periodic type, illustrating in detail what is meant by cyclic phenomena and how they are employed in solution.

SECTION IX.

SOLUTION OF PERIODIC POLYALPHABET CIPHERS USING STANDARD ALPHABETS.

	Paragraph
Classification.....	35
Method of encipherment—multiple alphabet ciphers.....	36
Steps in analysis—multiple alphabet ciphers.....	37
First step.....	38
Second step.....	39
Illustration of the solution of a multiple alphabet cipher using standard alphabets.....	40

35. Classification.—Ciphers of the periodic type may in turn be classified into two kinds: (1) Multiple alphabet ciphers, and (2) progressive alphabet ciphers. In the first, only a few of a whole series of cipher alphabets of which the system is composed may be used in a single message, and these alphabets repeat themselves in a definite sequence throughout the message. In the second, all of the cipher alphabets of the system are used in a single message, one after the other progressively, until the last alphabet has been used, when the series begins to repeat itself. The differences between the two systems will become more apparent in the subsequent discussion. Let us first proceed to a study of the common types of multiple alphabet ciphers.

36. Method of encipherment—multiple alphabet ciphers.—These systems usually employ a key consisting of a word or a number which determines the number of and the particular cipher alphabets to be used. This key is written on one line, and the text of the message is written in successive lines below it, thus forming columns of letters. Sufficient space is left between the lines of plain text for

the insertion of cipher equivalents. (See *a*, Fig. 11.) Then each column is enciphered by the cipher alphabet indicated by the key letter or key number at the head of the column, and the resulting lines of cipher text are sent in regular groups of five letters, as usual. An example employing standard alphabets is shown in Figure 11, using the key word BLUE.

This, in general, is the fundamental method employed in a multiple alphabet cipher. Manifold modifications in minor details are encountered. The number of alphabets employed in one message may vary from two to thirty or more, but as a general rule, in practice, the number is limited to an average of about ten. The alphabets may be of various types, too, but the general method of analysis is the same throughout.

37. Steps in analysis — multiple alphabet ciphers.—The analysis of a cryptogram of this type, regardless of the kind of cipher alphabets employed, whether they be standard or mixed alphabets, resolves itself into two distinct steps: (1) A determination of the exact number of cipher alphabets employed, followed by (2) an analysis of the individual cipher alphabets.

38. First step.—(a) *The principles of "factoring."*—In a cryptogram enciphered by a repeating key the determination of the number of alphabets involved is usually a simple matter, because the cryptogram itself affords clues to the length of the key. Why this is the case and the nature of these clues will now be explained.

It is obvious that identical plain-text letters when enciphered by the same alphabet must of necessity yield identical cipher letters. Such a condition is brought about every time that identical letters happen to fall in the same relative position as regards the key, or, if we refer to the diagram in Figure 11, every time identical letters fall within the same columns. Now, since the number of columns, or positions with respect to the key, is very limited (except in the case of very long key words), and since the repetition of letters is an inevitable condition in plain text, it follows that there will be in a message of fair length many cases where identical plain-text letters must fall in the same column and thus be enciphered by the same cipher alphabet, resulting, therefore, in the production of many identical letters in the cipher text.

It will also happen, however, that *different* plain-text letters falling in *different* columns will by mere coincidence produce identical cipher letters. Note, for example, in Figure 11 that in Column 1, R_p becomes S_o and that in Column 2, H_p also becomes S_o. The production of an identical cipher letter in these two cases, where the plain-text letters are different and enciphered by different alphabets, is merely a coincidence. Such cases will, of course, happen very frequently with

MESSAGE.

The artillery battalion marching in the rear of the advance guard keeps its combat train with it.

CIPHER ALPHABETS.

Plain-ABCDEFGHIJKLMNOPQRSTVWXYZ
 Cipher { (1)-BCDEFGHIJKLMNOPQRSTVWXYZA
 (2)-LMNOPQRSTUVWXYZABCDEFGHIJK
 (3)-UVWXYZABCDEFGHIJKLMNOPQRST
 (4)-EFGHIJKLMNOPQRSTVWXYZABCD

<u>B L U E</u>	<u>B L U E</u>	<u>B L U E</u>	<u>B L U E</u>
T H E A	T H E A	T H E A	T H E A
	U S Y E		U S Y E
R T I L	R T I L	D V A N	D V A N
	S E C P		E G U R
L E R Y	L E R Y	C E G U	C E G U
	M P L C		D P A Y
B A T T	B A T T	A R D K	A R D K
	C L N X		B C X O
A L I O	A L I O	E E P S	E E P S
	B W C S		F P J W
N M A R	N M A R	I T S C	I T S C
	O X U V		J E M G
C H I N	C H I N	O M B A	O M B A
	D S C R		P X V E
G I N T	G I N T	T T R A	T T R A
	H T H X		U E L E
H E R E	H E R E	I N W I	I N W I
	I P L I		J Y Q M
A R O F	A R O F	T H I T	T H I T
	B C I J		U S C X
<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>

CRYPTOGRAM.

USYES ECPMP LCCLN XBWCS OXUVD SCRHT HXIPL
 IBCIJ USYEE GURDP AYBCX OFPJW JEMGP XVEUE
 LEJYQ MUSCX

FIGURE 11.

individual letters, but less frequently with digraphs, because the chances that such a purely fortuitous coincidence will happen two times in succession are naturally much less than that it will happen every once in a while in the case of single letters; and, of course, the chances of such coincidences happening in the case of three or more consecutive letters are still less than in the case of digraphs. If we reduce the chances of such repetitions being due to pure coincidence to the form of a table, we have the following:

Chances for repetition of—

- 1 letter = 1:26 (i. e., 1 for, to 26 against).
- 2 letters = 1:676.
- 3 letters = 1:17,576.
- 4 letters = 1:456,976.
- n letters = 1:26ⁿ.

That an event the chances of whose occurrence is only one in 456,976 cases should be a coincidence is, of course, rarely to be expected, and it follows therefore that if in a relatively small amount of text we find a recurrence of a polygraph of four letters, for example, the chances that this recurrence is due to a mere coincidence are so remote that they may be altogether ruled out. If the recurrence is not a coincidence, then it must be due to a cause, and that cause is, of course, that a repeated polygraph in the plain text has been enciphered by similar alphabets. In order for this to occur, it is necessary that the polygraph fall both times in exactly the same relative position with respect to the key. Note, for example, the four letter repeated polygraph USYE in the message of Figure 11; it represents the plain-text polygraph THEA. The first time it occurred it fell in positions 1-2-3-4 with respect to the key; the second time it occurred it happened to fall in the very same relative positions, although it might just as well have happened to fall in any of the other three possible relative positions with respect to the key, viz, 2-3-4-1, 3-4-1-2, or 4-1-2-3. In fact, the word "happened" correctly expresses the case, for the insertion or deletion of a single plain-text letter between the two occurrences would have thrown the second occurrence one letter forward or backward, respectively, and thus caused the polygraph to be enciphered by a sequence of alphabets such as can no longer produce the cipher polygraph USYE from the plain-text polygraph THEA.¹ If, now, we count the number of letters *from and including the first occurrence* of USYE to *but not including the second occurrence* of USYE, we find a total of 40 letters. How many times has the key repeated itself between these two occurrences? Evidently $40 \div 4$, or 10 times. The number 40—that

¹ On the other hand, the insertion or deletion of this one letter might bring the letters of some other polygraph into similar columns so that another repetition would be exhibited in case the USYE repetition had thus been suppressed.

is, the interval separating the two occurrences—is of necessity an exact multiple of the length of the key. It is apparent, therefore, that if we did not know the length of the key the number 40, we would feel sure, must be an exact multiple of the length of the key; in other words, one of the factors of the number 40 would be equal to the length of the key. The factors of 40 are 2, 4, 5, 8, 10, and 20. So far as this single repetition, USYE, is concerned, the length of the key may be equal to any one of these factors; the repetition itself gives no indication. How, then, can we determine which one it is?

(b) *Application of the theory of factoring.*—Let us list all the recurrences in the cryptogram. They are as follows (note underlinings in the message):

USYES ECPMP LCCLN XBWCS OXUVD SCRHT
HXIPL IBCIJ USYEE GURDP AYBCX OFFJW
 JEMGP XVEUE LEJYQ MUSCX

1st USYE to 2nd USYE = 40 letters. Factors = 2, 4, 5, 8, 10, and 20.
 2nd US to 3rd US = 36 letters. Factors = 2, 3, 4, 6, 9, and 18.
 1st SC to 2nd SC = 52 letters. Factors = 2, 4, 13, and 26.
 1st PL to 2nd PL = 24 letters. Factors = 2, 3, 4, 8, 12.
 1st BC to 2nd BC = 16 letters. Factors = 2, 4, 8.
 1st CX to 2nd CX = 25 letters. Factor = 5.

What factors are the most common, i. e., the most frequently repeated, in this series?

The factors 2 and 4 each appear 5 times.

The factor 8 appears 3 times.

The factors 3 and 5 appear 2 times.

The factors 6, 9, 10, 12, 18, 20, and 26 each appear 1 time.

The most common factors are 2 and 4. Which of these two is the more probable as regards its being used as a key? Evidently 4, and we therefore assume the length of the key to be four letters, which means that four alphabets are involved. The recurrence of digraph CX is a pure coincidence, as will be apparent by referring to Figure 11. Had the message been longer there would have been more such pure coincidences, but, on the other hand, the chances are that there would also have been a proportionately greater number of real repetitions, for the greater the volume of text, the more repetitions of high-frequency digraphs, trigraphs, and polygraphs appear.

Sometimes it happens that the eye of the cryptanalyst immediately notes a repetition of a polygraph of four or more letters, the interval between the first and second occurrences of which has only two

factors, of which one is a relatively small number, the other a relatively high incommensurable number. He may therefore assume at once that the length of the key is equal to the smaller factor without searching for additional recurrences upon which to corroborate his assumption. Suppose, for example, that in a relatively short cryptogram the interval between the first and second occurrences of a polygraph of five letters happens to be a number such as 203, the factors of which are 7 and 29. Evidently the number of alphabets may at once be assumed to be 7, unless we are dealing with messages in which the correspondents are known to use long keys. In the latter case we could assume the number of alphabets to be 29.

In this process of factoring the greatest reliance is to be placed upon the longest repeated polygraphs. The evidence afforded by one repeated polygraph of four letters is very much greater than that afforded by many repeated digraphs, being in the ratio of 456, 976 to 676, or 676 to 1.

(c) *General remarks on factoring.*—The statement made in Par. 34 with respect to the cyclic phenomena said to be exhibited in cryptograms of the periodic type now becomes clear. The use of a short repeating key produces a periodicity of recurrences or repetitions collectively termed "cyclic phenomena," an analysis of which leads to a determination of the length of the period or cycle, and this gives us the length of the key. Only in the case of relatively short cryptograms enciphered by a relatively long key does this process of factoring fail to lead to the correct determination of the number of cipher alphabets in a multiple alphabet cipher, and, of course, the fact that a cryptogram contains repetitions whose factors show constancy is in itself an indication and test of its multiple alphabet nature. It also follows that if the cryptogram is not a multiple alphabet cipher, then the process of factoring will show no definite results, and conversely the fact that it does not yield definite results at once indicates that the cryptogram is not a multiple alphabet cipher. There are two cases in which the process of factoring leads to no definite results. One is in the case of monoalphabet substitution ciphers. Here recurrences are very plentiful as a rule, and the intervals separating these recurrences may be factored, *but the factors will show no constancy*; there will be several factors common to many or most of the recurrences. This in itself is an indication of a monoalphabet cipher, if the very fact of the presence of many recurrences fails to impress itself upon the inexperienced cryptanalyst. The other case in which the process of factoring is non-significant involves certain types of nonperiodic polyalphabet ciphers. In certain of these ciphers recurrences of digraphs, tri-graphs, and even polygraphs may be plentiful in a long message,

but the intervals between such recurrences bear no definite multiple relation to the length of the key, such as in the case of the true periodic multiple alphabet cipher, in which the alphabets change with successive letters and repeat themselves over and over again.

89. Second step.—(a) *Preparation of individual graphic frequency tables.*—After the number of cipher alphabets involved in the cryptogram has been ascertained by factoring, the next step is to rewrite the message in groups corresponding to the length of the key, or in columnar fashion, whichever is more convenient, and this automatically divides up the text so that the letters belonging to the same cipher alphabet occupy similar positions in the groups, or, if the columnar method is used, fall in the same column. Then single frequency tables for the thus isolated individual alphabets are compiled. For example, in the case of the cipher on page 41, having determined that four alphabets are involved, and having rewritten the message into four columns, a frequency table is made of the letters in Column 1, another frequency table is made of the letters in Column 2, and so on for the rest of the columns. *Each of the resulting tables is therefore a monoalphabet frequency distribution.* If these tables do not give the irregular crest and trough appearance of single frequency tables, then the analysis which led to the hypothesis as regards the number of alphabets involved is fallacious. In fact, the appearance of these individual graphic tables may be considered to be an index of the correctness of the factoring process; for theoretically, and practically, the individual graphic tables constructed upon the *correct* hypothesis will tend to conform more closely to the irregular crest and trough appearance of a single alphabet frequency distribution than will the graphic tables constructed upon an incorrect hypothesis. For example, in the following cryptogram there are only two cases of repetitions of three or more letters, and the factors of the intervals separating them do not show conclusively whether there are five or six alphabets involved:

QNDRV AAALX POAQR XPFQN QZHRK ZVPLO JPJUK
 PMEKV TNUKN JLV RN KVTNI AHDRV AAWKR EAAWO
 APTHK WQXKO YNDGJ YAJUV RXPFI RKPQI TPFUV
 YIRZF TZHUN YIBJI GPAMW UDBJO YZBY

1st DRVAA to 2d DRVAA = 55 letters; factors 5, 11.

1st KVTN to 2d KVTN = 12 letters; factors 2, 3, 4, 6.

1st AAW to 2d AAW = 6 letters; factors 2, 3, 6.

But if we make a graphic frequency table for the first alphabet upon each of the two more probable hypotheses, we see that the distribution upon the basis of five alphabets is more favorable to that

hypothesis than is the distribution upon the basis of six alphabets to a hypothesis of six alphabets.

Alphabet 1 upon the basis of five alphabets:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 ≡

Alphabet 1 upon the basis of six alphabets:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 ≡

In the first distribution there are only seven cases in which a letter occurs but once; in the second, there are ten such cases; in the first, there are four cases in which a letter occurs twice; in the second, there are only three such cases; in the first, there are three letters which occur three or more times; in the second, there are only two such cases. In other words, the principal characteristics of a single frequency distribution are more closely approximated by the distribution upon the hypothesis of five alphabets, as compared with that upon the hypothesis of six alphabets. The greater the degree of repetition in a frequency table, the more likely is it to be a single frequency distribution.

(b) *Analysis of the graphic tables.*—The difficulty experienced in analyzing the individual or isolated graphic tables depends mostly upon the type of cipher alphabets that is used. It is apparent that mixed alphabets may be used just as easily as standard alphabets, and, of course, the cipher letters themselves give no indication as to which is the case. However, just as we found that in the case of monoalphabet substitution ciphers a graphic frequency table will give clear indications whether the cipher alphabet is a standard or a mixed alphabet, by the relative positions and extensions of the crests and troughs in the table, so we find that in the case of periodic multiple alphabet substitution ciphers, graphic frequency tables for the isolated or individual alphabets will also give clear indications as to whether these alphabets are standard alphabets or mixed alphabets. Only one or two graphic frequency tables are necessary for this determination; if they appear to be standard alphabets, graphic tables can be made for the rest of the alphabets; but if they appear to be mixed alphabets, then it is best to compile trigraphic frequency tables for all the alphabets. The analysis of the values of the cipher letters in each table proceeds along the same lines as in the case of monoalphabet ciphers. The analysis is more difficult only because of the reduced size of the tables, but if the message be very long, then each frequency table will contain a sufficient number of elements to enable a speedy solution to be achieved.

40. Illustration of the solution of a multiple alphabet cipher using standard alphabets.—(a) *Frequency table method.*—In the light of the foregoing principles let us study the following cryptogram:

MESSAGE.

AUKHY JAMKI ZYMWM JMIGX NFMLX ETIMI ZHBHR
 AYMZM ILVME JKUTG DPVXK QUKHQ LHVRM JAZNG
 GZVXE NIUFM PZJNV CHUAS HKQGK IPLWP AJZXI
 GUMTV DPTEJ ECMYS QYBAV ALAHY POEXW PVNYE
 EYXEE UDPXR BVZVI ZIIVO SPTEG KUBBR QLLXP
 WFQGK NLLLE PTIKW DJZXI GOIOI ZLAMV KFMWF
 NPLZI OVVF M ZKTXG NLMDF AAEXI JLUFM PZJNV
 CAIGI UAWPR NVIWE JKZAS ZLAFM HS

A search for repetitions discloses the following short list, with the intervals and factors below 11 listed (for previous experience may lead us to suspect that it is unlikely that the cryptogram involves more than 10 alphabets, showing the number of recurrences it does):

UFMPZJNV	—160 = 2, 4, 5, 8, 10.	FM	—20 = 2, 4, 5.
JZXIG	—90 = 2, 3, 5, 6, 9, 10.	FM	—30 = 2, 3, 5.
QGK	—85 = 5.	JA	—60 = 2, 3, 4, 5, 6, 10.
EJK	—315 = 3, 5, 7.	LA	—75 = 3, 5.
UKH	—55 = 5.	LL	—10 = 2, 5.
ZLA	—65 = 5.	NL	—45 = 3, 5, 9.
EJ	—69 = 3.	NL	—105 = 3, 5, 7.
FM	—185 = 5.	VX	—20 = 2, 4, 5, 10.
FM	—57 = 3.	YM	—25 = 5.
FM	—140 = 2, 4, 5, 7.		

The factor 5 appears in all but two cases, both of which exceptions involve only a digraph, and we may feel certain that the number of alphabets is five. Since the text already appears in groups of five letters, it is unnecessary to rewrite the message. We next proceed to make a graphic frequency table for Alphabet 1 to see if we can determine whether or not standard alphabets are involved. It is as follows:

Alphabet 1.

X	=	≡	≡	≡	≡	X	=	X	≡	X	≡	=	=	X	=	X									
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Although the indications are not very clear cut, yet if we take into consideration the small size of the table the assumption that we have here a direct standard alphabet with $W_o = A_p$, is worth further test. Let us compile a similar table for Alphabet 2.

Alphabet 2.

$\begin{array}{cccccccccccccccccccc}
\text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} & \text{I} & \text{J} & \text{K} & \text{L} & \text{M} & \text{N} & \text{O} & \text{P} & \text{Q} & \text{R} & \text{S} & \text{T} & \text{U} & \text{V} & \text{W} & \text{X} & \text{Y} & \text{Z} \\
\text{---} & \text{---}
\end{array}$

Here we have every indication of a direct standard alphabet, with $H_o = A_p$. Let us make similar tables for the last three alphabets. They are as follows:

Alphabet 3.

$\begin{array}{cccccccccccccccccccc}
\text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} & \text{I} & \text{J} & \text{K} & \text{L} & \text{M} & \text{N} & \text{O} & \text{P} & \text{Q} & \text{R} & \text{S} & \text{T} & \text{U} & \text{V} & \text{W} & \text{X} & \text{Y} & \text{Z} \\
\text{---} & \text{---}
\end{array}$

Alphabet 4.

$\begin{array}{cccccccccccccccccccc}
\text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} & \text{I} & \text{J} & \text{K} & \text{L} & \text{M} & \text{N} & \text{O} & \text{P} & \text{Q} & \text{R} & \text{S} & \text{T} & \text{U} & \text{V} & \text{W} & \text{X} & \text{Y} & \text{Z} \\
\text{---} & \text{---}
\end{array}$

Alphabet 5.

$\begin{array}{cccccccccccccccccccc}
\text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} & \text{I} & \text{J} & \text{K} & \text{L} & \text{M} & \text{N} & \text{O} & \text{P} & \text{Q} & \text{R} & \text{S} & \text{T} & \text{U} & \text{V} & \text{W} & \text{X} & \text{Y} & \text{Z} \\
\text{---} & \text{---}
\end{array}$

After but little experiment we find that the graphic tables can best be made to fit the normal table when the following values are assumed:

Alphabet 1— $A_p = W_o$.

Alphabet 2— $A_p = H_o$.

Alphabet 3— $A_p = I_o$.

Alphabet 4— $A_p = T_o$.

Alphabet 5— $A_p = E_o$.

Note the key word given by the successive equivalents of A_p : WHITE. The real proof of the correctness of our analysis is, of course, to test the values of the solved alphabets on the cryptogram. The five complete cipher alphabets are as follows:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
Cipher	1—	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	2—	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
	3—	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	4—	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	5—	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

Applying these values to the first few groups of our message, we have the following:

	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
Cipher—	AUKHY	JAMKI	ZYMWM	JMIGX	NFMLX	. . .
Plain—	ENCOU	NTERE	DREDI	NFANT	RYEST	. . .

Intelligible text at once results, and the solution can now be completed very quickly. The message is as follows:

Encountered red infantry estimated at one regiment and machine gun company in trucks near EMMITSBURG. Am holding MIDDLE CREEK near hill 543 southwest of FAIRPLAY. When forced back will continue delaying Reds at MARSH CREEK. Have destroyed bridges on MIDDLE CREEK between EMMITSBURG-TANEYTOWN road and RHODES MILL.

It is obvious that reversed standard alphabets may be used and treated in the same manner. In fact, the now obsolete cipher disk used by the United States Army for a number of years yields exactly this type of cipher and may just as readily be solved. In fitting the isolated graphic tables to the normal the direction of "reading" the crests and troughs is merely reversed.

(b) *Completion of plain component method.*—There is another method of solving this type of cipher, which is worth while explaining, because the underlying principles will be found exceedingly useful in many cases.

After all, the individual alphabets of a cipher such as the one just solved are merely standard direct alphabets. We have seen that in the case of monoalphabet ciphers in which standard cipher alphabets are employed they may be solved almost mechanically by completing the plain component sequence (see page 16). The plain text reappears on only one generatrix, and this generatrix is the same for the whole message. We were able to pick this generatrix out of all the other generatrices because it was the only one which gave intelligible text. Is it not apparent that if we apply the same process to the cipher letters of the *individual alphabets* of the cipher just solved that the plain-text equivalents of these letters must all reappear on one and the same generatrix? But how can we pick out the correct generatrix from among all the other incorrect generatrices, since the correct one will not show intelligible text? The answer is simple. We should be able to select it *because it will show more and a better assortment of high-frequency letters than any of the other generatrices*. If we do this with all the alphabets in the cryptogram, it will merely be necessary to assemble the correct generatrices in proper order, and the result should be intelligible text. An example will serve to make the process clear. Let us use the same message as before. Factoring

showed that it involves five alphabets. Let us set down in a horizontal line the first ten cipher letters in each alphabet and complete the normal alphabet sequences. Thus:

	Alphabet 1.	Alphabet 2.	Alphabet 3.	Alphabet 4.	Alphabet 5.
1	<u>AJZJNEZAIJ</u>	UAYMFTHYLK	KMMIMIBMVU	HKWGLMHZMT	YIMXXIRMEG
2	BKAKOFABJK	VBZNGUIZML	LNNJNJCNWV	ILXHMNIANU	ZJNYYJSNFH
3	CLBLPGBCKL	WCAOHVJANM	MOOKOKDOXW	JMYINOJBOV	AKOZZKTOGI
4	DMCMQHCGLM	XDBPIWKBN	NPPLPLEPYX	KNZJOPKCPW	BLPAALUPHJ
5	<u>ENDNRIDEMN</u>	YECQJXLCPO	OQQMQMFQZY	LOAKPQLDQX	CMQBBMVQIK
6	FOEOSJEFNO	ZFDRKYMDQP	PRRNRNGRAZ	MPBLQRMERY	DNRCCNWRJL
7	GPFPTKFGOP	AGESLZNERQ	QSSOSOHSBA	NQCMRSNFSZ	EOSDDOXSKM
8	HQQQULGHPQ	BHFTMAOFSR	RTTPTPITCB	<u>ORDNSTOCTA</u>	FPTEEPYTLN
9	IRHRVMHIQR	CIGUNBPGTS	SUUQUQJUDC	PSEOTUPHUB	GQUFFQZUMO
10	JSISWNIJRS	DJHVOCQHUT	TVVRVRKVED	QTFPUVQIVC	HRVGGRAVNP
11	KTJTXOJKST	EKIWPDRIVU	UWWSWSLWFE	RUGQVWRJWD	ISWHHSBWOQ
12	LUKUYPKLTU	FLJXQESJWV	VXXTXTMXGF	SVHRWXSKE	JTXIITCXPR
13	MVLVZQLMUV	GMKYRFTKXW	WYYUYUNYHG	TWISXYTYLF	KUYJJUDYQS
14	NMWARMNVW	HNLZSGULYX	XZZVZVOZIH	UXJTYZUMZG	LVZKKVEZRT
15	OXNXBSNOWX	IOMATHVMZY	YAAWAWPAJI	VYKUZAVNAH	MWALLWFASU
16	PYOYCTOPXY	JPNBUIWNAZ	ZBBXBQBKJ	WZLVABWOBI	NXBMMXGBTV
17	QZPZDUPQYZ	KQOCVJXOBA	ACCYCYRCLK	XAMWBCXPCJ	OYCNNYHCUW
18	RAQAEVQRZA	LRPDWKYPCB	BDDZDZSDML	YBNXCDYQDK	PZD00ZIDVX
19	SBRBFWRSAB	MSQEXLZQDC	<u>CEEAEATENM</u>	ZCOYZEZREL	QAEPJAJEWY
20	TCSGXSTBC	<u>NTRFYMARE</u>	DFFBFBUFON	ADPZEFASFM	RBFQQBKFXZ
21	UDTDHYTUCD	OUSGZNBSFE	EGGCGCVGPO	BEQAFGBTGN	SCGRRCLGYA
22	VEUEIZUVDE	PVTHAOCYGF	FHHDHDWHQP	CFRBGHCUHO	TDHSSDMHZB
23	WVVFJAVWEF	QWUIBPDUGH	GIIEIEXIRQ	DGSCHIDVIP	<u>UEITTENIAC</u>
24	XGWGKBWYFG	RXVJCQEVII	HJJFJFYJSR	EHTDIJEWJQ	VFJUUFQJBD
25	YHXHLCXYGH	SYWKDRFWJI	IKKKGZKTS	FIUEJKFKKR	WGKVVGPKE
26	ZIYIMDYZHI	TZXLESGXKJ	JLLHLHALUT	GJVFKLGYLS	XHLWWHQLD

FIGURE 12.

If now we select the following high-frequency generatrices (underlined in Fig. 12):

For Alphabet 1, generatrix 5—E N D N R I D E M N

For Alphabet 2, generatrix 20—N T R F Y M A R E D

For Alphabet 3, generatrix 19—C E E A E A T E N M

For Alphabet 4, generatrix 8—O R D N S T O G T A

For Alphabet 5, generatrix 23—U E I T T E N I A C

and arrange their letters in *columnar* order, thus:

E N C O U
 N T E R E
 D R E D I
 N F A N T
 R Y E S T
 I M A T E
 D A T O N
 E R E G I
 M E N T A
 N D M A C

we have intelligible text: ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MAC Solution can thus be achieved without the compilation of any frequency tables whatever, and is very quickly attained. The inexperienced cryptanalyst may have difficulty at first in selecting the generatrix which contains the most and the best assortment of high-frequency letters, but with increased practice, a high degree of proficiency is attained. After all it is only a matter of experiment, trial, and error to select and assemble the proper generatrices so as to produce intelligible text. If the letters on our sliding strips were accompanied by numbers proportionate to their frequency in normal plain text, then that generatrix, the relative frequency values of whose letters totaled the greatest, would *theoretically* always be the correct generatrix. Practically it will be among the generatrices which show the first three or four greatest totals. Thus, an entirely mathematical solution for this type of cipher may be applied.

If the cipher alphabets are reversed standard alphabets, it is only necessary to convert the cipher letters of each isolated alphabet into their normal plain component equivalents and then proceed as in the case of direct standard alphabets.

We have seen how the key word may be discovered in this type of cryptogram. Usually the key is made up of those letters in the successive alphabets whose equivalents are A_p . Sometimes a key number is used, such as 8-4-7-1-12, which means merely that A_p is represented by the eighth letter from A (in the normal alphabet) in the first cipher alphabet, by the fourth letter from A in the second cipher alphabet, and so on. However, the method of solution as illustrated above, being independent of the nature of the key, is the same as before.

(c) *Reconstruction of key method.*—The common use of key words in such cryptograms makes possible a method of solution that is simple and can be used where the more detailed method of analysis by frequency tables or by completing the plain component is of no avail, so that in the case of a very short message which may

show no recurrences and give no indications as to the number of alphabets involved, this modified method will be found useful.

Briefly, the method consists in assuming the presence of a probable word in the message, and referring to the alphabets to find the key letters concerned when this hypothetical word is assumed to be present in various positions in the cipher text. If the assumed word happens to be correct, and is placed in the correct location in the message, the key letters produced by referring to the alphabets will yield the key word. In the following example it is assumed that reversed standard alphabets are known to be used by the enemy.

MESSAGE.

MDSTJ LQCXC KZASA NYKOLP

Extraneous circumstances lead us to assume the presence of the word AMMUNITION. We begin by assuming that this word begins the message. Using sliding normal alphabets, one reversed, the other direct, we proceed to find the key letters by noting what the successive equivalents of A_p are. Thus:

If M D S T J L Q C X C equals
A M M U N I T I O N, then the key letters ($=A_p$) are
M P E N W T J K L P.

The key does not spell any intelligible word. We therefore shift our assumed word one letter forward and try again.

If D S T J L Q C X C K equals
A M M U N I T I O N, then the key letters ($=A_p$) are
D E F D Y Y V F Q X.

This neither yields an intelligible key word. We shift the assumed word forward one space at a time until we strike the following point:

If L Q C X C K Z A S A equals
A M M U N I T I O N, then the key letters ($=A_p$) are
L C O R P S S I G N.

The key stands out: SIGNAL CORPS.¹ If the assumption of reversed standard alphabets yields no good results, then direct standard alphabets are assumed and the test made exactly in the same manner. Solution by this method is inevitable when the correct word has been assumed and its correct position ascertained. This method, as will be shown subsequently, can also be used as a last resort when mixed alphabets are employed.

¹ It should be clear that since the key word or key phrase repeats itself during the encipherment of such a message, the plain-text word upon whose assumed presence in the message this test is being based may begin to be enciphered at any point in the key, and continue over into its next repetition. When this is the case it is merely necessary to shift the latter part of the sequence of determined key letters to the first part, as in the case noted: LCORPSSIGN is transposed into SIGNLCORPS, and thus SIGNAL CORPS.

SECTION X.

SOLUTION OF PERIODIC POLYALPHABET CIPHERS USING MIXED ALPHABETS.

CASE I.—*The plain component is the normal sequence.*

	Paragraph.
Reason for the use of mixed alphabets.....	41
Interrelated mixed alphabets as produced by sliding components.....	42
Reconstructing the cipher component; the principles of symmetry of position.....	43
Application of foregoing principles—solution of original messages.....	44
Results of factoring.....	44a
Examination of frequency tables.....	44b
Results of deductions.....	44c
Application of principles of symmetry.....	44d
Substitution of deduced values.....	44e
Completing the solution.....	44f
Application of foregoing principles—solution of subsequent messages.....	45
Factoring and conversion into plain component equivalents.....	45a
Examination and selection of generatrices.....	45b
Combining the selected generatrices.....	45c
Application of foregoing principles—solution by reconstruction of key.....	46

41. Reason for the use of mixed alphabets.—We have seen in the examples considered thus far that the use of several alphabets in the same message does not greatly complicate the analysis of such a cryptogram. There are three reasons why this is so: Firstly, only a few alphabets were employed; secondly, these alphabets were employed in a periodic or repeating manner, giving rise to cyclic phenomena in the cryptogram, by means of which the number of alphabets could be determined; and, thirdly, the cipher alphabets were *known* alphabets, by which is meant merely that the sequences of letters in both components of the cipher alphabets were known sequences. We shall now consider the effects of modifying the last of these factors of the analysis.

In the case of monoalphabet ciphers we found that the use of a mixed alphabet delayed the solution to a considerable degree, and we shall now see that the use of mixed alphabets in polyalphabet ciphers renders the analysis much more difficult than the use of standard alphabets, but the solution is still fairly easy to achieve.

42. Interrelated mixed alphabets as produced by sliding components.—It was stated in Par. 33 that the method of producing the mixed alphabets in a polyalphabet cipher often affords clues which are of great assistance in the analysis of the cipher alphabets. This is so, of course, only when the cipher alphabets are interrelated secondary alphabets produced by sliding components. The second type of components listed on page 35, viz, that in which the components are not both normal sequences, was further subdivided into two cases, the first of which, namely, when the plain component is the normal sequence, will now be considered.

Here one of the components, usually the plain component, is the normal sequence, while the other is a mixed sequence, the sliding of the two components yielding mixed alphabets. The mixed com-

ponent may be a systematically mixed or a random-mixed sequence, and very often the systematically mixed sequence consists of a key-word sequence (see page 29). If we set down the successive alphabets produced by the sliding of two such components, as in the case of the Vigenère Table, we have a symmetrical table such as that shown in Table 3.

TABLE 3.

Plain.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	
E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	
A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	
V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	
N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	
W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	
O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	
R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	
T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	
H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	
B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	
C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	
D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	
F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	
G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	
I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	
J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	
K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	
M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	
P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	
Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	
S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	
U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	
X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	
Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	
Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	

Cipher

Such a table may be used in exactly the same manner as the Vigenère Table. With the key word BLUE the following secondary alphabets would be used:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
2	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
3	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
4	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

Cipher

43. Reconstructing the cipher component; the principles of symmetry of position.—It was stated directly above that Table 3 is a symmetrical table, by which is meant that the letters in its successive horizontal lines show a *symmetry of position* with respect to one another. They constitute, really, one and only one sequence or series of letters, and this fact can be used to good advantage. Consider, for example, the pair of letters G and V in the B, or 1st, cipher alphabet directly above; the letter V is the 15th letter to the right of G. In the L, or 2d, cipher alphabet, V is also the 15th letter to the right of G, as is the case in every one of these secondary alphabets, since the *relative* positions they occupy are the same in each horizontal line, that is, in each cipher alphabet. If, therefore, we know the relative positions occupied by a given pair of letters in one of these cipher alphabets, and have located one of the members of this same pair in another of these cipher alphabets, we may at once place the other member of this pair in its proper position in the second of the cipher alphabets. Suppose, for example, that as the result of an analysis based upon considerations of frequency, we have assumed the following values in a given cryptogram:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1—					G										Y					V						
2—					N										G					P						
3—					L										B					I						
4—					W										I					Q						

FIGURE 13.

The letter G is common to Alphabets 1 and 2. In Alphabet 2 we note that N occupies the 10th position to the left of G, and the letter P occupies the 5th position to the right of G. We may therefore place these letters, N and P, in their proper positions in Alphabet 1, the letter N being placed 10 letters before G, and the letter P, 5 letters after G. Thus:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 1— G P Y V N

Thus we have the values of two new letters in Alphabet 1, viz, $P_o = J_p$, and $N_o = U_p$; these values were obtained without any analysis based upon the *frequency* of P_o and N_o .

Likewise, in Alphabet 2, we may insert the letters Y and V in these positions:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 2— V N G P Y

This gives the new values $V_c = D_p$ and $Y_c = Y_p$ in Alphabet 2.

Alphabets 3 and 4 have a common letter I, which permits of the placement of Q and W in Alphabet 3, and of B and L in Alphabet 4. The new values thus found are of course immediately inserted throughout the cryptogram, thus leading to the assumption of further values in the cipher text. This process, the reconstruction of the primary components by the application of the principles of symmetry of position, thus facilitates and hastens solution.

44. Application of foregoing principles—solution of original messages.—In the light of the foregoing principles let us now study a typical message.

MESSAGE.

SIJYU MNVCA ISPJL RBZEY QWYEU LWMGW ICJCI MTZEI MIBKN
 QWBRI VWYIG BWNBQ QCGQH IWJKA GEGXN IDMRU VEZYG QIGVN
 CTGYO BPDFL VCGXG BKZZG IVXCU NTZAO BWFEO QLFEO MTYZT
 CCBYQ OPDKA GDGIG VPWMR QIIIEW ICGXG BLGQQ VBGRS MYJJY
 QVFWY RWNFL GXNFW MCJKX IDDRU OPJQQ ZRHCN VWDYQ RDGDG
 BXDBN PXPFU YXNFG MPJEL SANCD SEZZG IBEYU KDHCA MBJJF
 KILCJ MFDZT CTJRD MIYZQ ACJRR SBGZN QYAHQ VEDCQ LXNCL
 LVVCS QWBII IVJRN WNBRI VPJEL TAGDN IRGQP ATYEW CBYZT
 EVGQU VPYHL LRZNQ XINBA IKWJQ RDZYP KWFZL GWFJQ QWJYQ
 IBWRX

44a. Results of factoring.—The principal repetitions of three or more letters have been underlined in the message and the factors (up to 20 only) of the intervals between them are as follows:

$$\text{CGXGB} - 60 = 2, 3, 4, 5, 6, 10, 12, 15, 20.$$

$$\text{PJEL} - 95 = 5, 19.$$

$$\text{BRI} - 285 = 3, 5, 15, 19.$$

$$\text{QRD} - 165 = 3, 5, 15.$$

$$\text{QWB} - 275 = 5, 11.$$

$$\text{WIC} - 130 = 2, 5, 10, 13.$$

$$\text{XNF} - 45 = 3, 5, 9, 15.$$

$$\text{YZT} - 225 = 3, 5, 15.$$

$$\text{ZGI} - 145 = 5.$$

The factor 5 is common to all of these repetitions, and there seems to be every indication that five alphabets are involved. Since the message already appears in groups of five letters, it is unnecessary in this case to rewrite it in groups corresponding to the length of

the key. We proceed at once to make a graphic frequency table for Alphabet 1. It is as follows:



Attempts to fit this table to the normal on the basis of a direct or reversed standard alphabet do not give positive results, and we assume that mixed alphabets are involved. Individual *trigraphic* frequency tables are then compiled and are shown in Figure 14. These tables are similar to those made for single mixed alphabet ciphers, and are made in the same way except that instead of taking the letters one after the other, we now must assemble in separate tables the letters which belong to the separate alphabets. For example, in Alphabet 1, the trigraph QAC means that A occurs in Alphabet 1; Q, its prefix, occurs in Alphabet 5, and C, its suffix, occurs in Alphabet 2. We may avoid all confusion by placing numbers indicating the alphabets in which they belong above the letters, thus: $\overset{5}{Q}\overset{1}{A}\overset{2}{C}$.

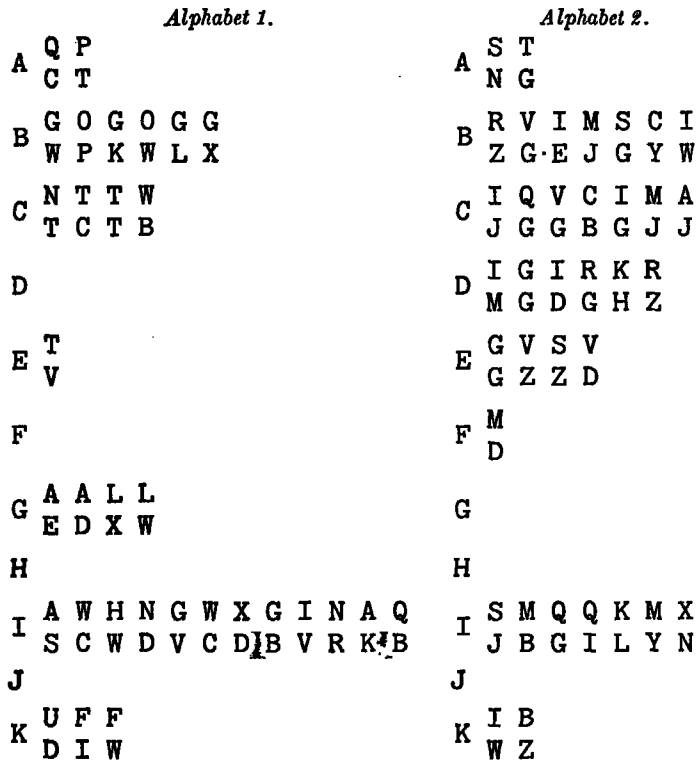


FIGURE 14.

ELEMENTS OF CRYPTANALYSIS.

Alphabet 1—Continued.

L U Q L L
 W X V R
 M U I I O S W G A J D
 N T I T Y C P B F I
 U
 T
 O Q U
 P P
 P N
 X
 Q Y N Q G Q R Y N S Q
 W W C I L I V Y W W
 R L Y Q Q
 B W D D
 S - L D R
 I A E B
 T L
 A
 U
 V I U L G Q N Q I U
 W E C P B W E P P
 W N
 N
 X Q
 I
 Y U
 X
 Z Q
 R

Alphabet 3.

A Y
 H
 B I W C W N
 K R Y I R
 C
 D P P D W X F E
 B K R Y B Z C
 E B
 Y

Alphabet 2—Continued.

L Q B
 F G
 M
 N M W
 V B
 O
 P B O V O M V V
 D D W J J J Y
 Q
 R Z I L
 H G Z
 S I
 P
 T M C N M C A
 Z G Z Y J Y
 U
 V I Q L I E
 X F V J G
 W Q L Q V B I B R V Q K G Q
 Y M B Y N J F N D B F F J
 X G B P Y L
 N D F N N
 Y M Q
 J A
 Z

Alphabet 4.

A Z
 O
 B N D D N
 Q L N A
 C V J X F H N H L D N V
 A I U O N D A J Q L S
 D G G
 G N
 E Z Y Z F I J J Y
 Y U I Q W L L W

FIGURE 14 (continued).

Alphabet 3—Continued.

F W L V X W W
E C W P Z J

G C E I T C D C L B D B A R V
Q X V Y X I X Q R D Z D Q Q

H R D
C C

I I
E

J I C W Y C P P B T C V P W
Y C K J K Q E J R R R E Y

K

L I
C

M W D
G R

N W W X X A X I
B F F F C C B

O

P S
J

Q

R

S

T

U

V N V
C C

W P K B
M J R

X V
C

Y W W T I T B P
E I Z Z E Z H

Z B T E K T E R D
E E Y Z A Z N Y

Alphabet 4—Continued.

F N N N
L W G

G M
W

H A Y
Q L

I Y G B
G G I

J P J J W F
L Y F Q Q

K B J D J
N A A X

L

M W
R

N Z
Q

O

P F
U

Q G G J G G
H Q Q P U

R B M G D J J J B W
I U S U D R N I X

S

T

U

V G
N

W F
Y

X G G G
N G G

Y J Z G B D E Z J
U G O Q Q U F Q

Z Z Y Z D Y G Y F
G T G T Q N T L

FIGURE 14 (continued).

58

ELEMENTS OF CRYPTANALYSIS.

Alphabet 5.

A C K K C B
 I G G M I
 B
 C
 D C R
 S M
 E
 F J Y
 K K
 G I Y X Z I X D F Z
 B Q B I V B B M I
 H Q
 I
 I C E R I R
 M M V I V
 J C
 M
 K
 J B F E C E H Z
 L R V G S L T L G
 M
 K X V C B Z R D
 N Q I C V P Q W I
 O Y A C
 B B M
 P Q
 A
 B E Y Q Q Y Z H C N J J Y
 Q Q Q O V Z R A V L X R Q I
 R M R
 Q S
 S R C
 M Q
 T Z Z Z
 C C E
 U Y E R C R P Y Q
 M L V N O Y K V
 V
 W G E F E
 I I M C
 X K R
 I -
 Y E J W
 Q Q R
 Z

Condensed table of repetitions.

5-1-2	1-2
W I C-2	Q W-4
Q R D-2	B W-3
	V P-3
	2-3
	C G-3
1-2-3	C J-3
Q W B-2	P J-3
	W F-3
	X N-3
	3-4
2-3-4	G Q-4
C G X-2	G X-3
P J E-2	J R-3
X N F-2	N F-3
	Y Z-3
	3-4-5
J E L-2	
B R I-2	4-5
G X G-2	Y Q-3
Y Z T-2	Z T-3
Z Z G-2	
	4-5-1
K A G-2	5-1
Z G I-2	G B-4
X G B-2	Q Q-3
Z T C-2	

44b. **Examination of frequency tables.**—We now proceed to analyze each alphabet. There seems to be no doubt about the equivalent of E_p in each alphabet; $E_p = \overset{1}{I}_c, \overset{2}{W}_c, \overset{3}{G}_c, \overset{4}{C}_c, \overset{5}{Q}_c$. Let us now see if we can separate the vowels from the consonants in each alphabet, using the same test as in the case of the mixed alphabet cipher on page 25.

The letters of greatest frequency in Alphabet 1 are I, M, Q, V, B, G, L, R, S, and C. I_c has already been assumed to be E_p . If $\overset{2}{W}_c$ and $\overset{5}{Q}_c = E_p$, then we should be able to determine the vowels from the consonants among the letters M, Q, V, B, G, L, R, S, and C by examining the prefixes of $\overset{2}{W}_c$, and the suffixes of $\overset{5}{Q}_c$. The prefixes and suffixes of these letters, as shown by the trigraphic frequency tables, are these:

Prefixes of $\overset{2}{W}_c (= \overset{2}{E}_p)$

Q G K V R B I L

Suffixes of $\overset{5}{Q}_c (= \overset{5}{E}_p)$

I Q R X L V A Z O

Consider now the letter $\overset{1}{M}_c$; it does not occur either as a prefix of $\overset{2}{W}_c$, or as a suffix of $\overset{5}{Q}_c$. Hence it is most probably a vowel, and on account of its high frequency it may be assumed to be O_p . On the other hand, note that $\overset{1}{Q}_c$ occurs four times¹ as a prefix of $\overset{2}{W}_c$ and three times as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, most probably R, for it would give the digraph ER ($= \overset{5}{Q}\overset{1}{Q}_c$) as occurring three times and RE ($= \overset{1}{Q}\overset{2}{W}_c$) as occurring four times.

The letter $\overset{1}{V}_c$ occurs twice as a prefix of $\overset{2}{W}_c$ and twice as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, and on account of its frequency, let us assume it to be T_p . The letter $\overset{1}{B}_c$ occurs twice as a prefix of $\overset{2}{W}_c$ but not as a suffix of $\overset{5}{Q}_c$. Its frequency is only medium, and it is probably a consonant. In fact, the thrice repeated digraph $\overset{1}{B}\overset{2}{W}_c$ is once a part of the trigraph $\overset{5}{G}\overset{1}{B}\overset{2}{W}$, and $\overset{5}{G}_c$, the letter of second highest frequency in Alphabet 5, looks excellent for T_p . Might not the trigraph $\overset{5}{G}\overset{1}{B}\overset{2}{W}$ be THE? Let us keep the possibility in mind.

The letter $\overset{1}{G}_c$ occurs only once as a prefix of $\overset{2}{W}_c$ and does not occur as a suffix of $\overset{5}{Q}_c$. It may be a vowel, but we can not be sure. The letter $\overset{1}{L}_c$ occurs once as a prefix of $\overset{2}{W}_c$ and once as a suffix of

¹ The letter Q has three tallies under it, plus one occurrence indicated by the presence of the letter itself among the prefixes, equals four occurrences. The same applies to the other letters.

Q_c^5 . We may consider it to be a consonant. R_c^1 occurs once as a prefix of W_c^2 , and twice as a suffix of Q_c^5 , and is certainly a consonant. Neither the letter S_c^1 nor the letter C_c^1 occurs as a prefix of W_c^2 nor as a suffix of Q_c^5 ; both would seem to be vowels, but a study of the prefixes and suffixes of these letters lends more weight to the assumption that C_c^1 is a vowel than that S_c^1 is a vowel. For all the prefixes of C, viz, N^5 , T^5 , and W^5 , are in subsequent analysis of Alphabet 5 classified as consonants, as are likewise its suffixes, viz, T^1 , C^1 , and B^1 in Alphabet 2. On the other hand, only one prefix, L_c^5 , and one suffix, B_c^2 , of S_c^1 are later classified as consonants. Since vowels are more often associated with consonants than with other vowels, it would seem that C_c^1 is more likely to be a vowel than S_c^1 . Let us at any rate assume C_c^1 to be a vowel, for the present leaving S_c^1 unclassified.

44c. Results of deductions.—When we go through the same steps with the remaining alphabets, we have the following results:

Alphabet.	Consonants.	Vowels.
1—	Q, V, B, L, R, G?	I, M, C
2—	B, C, D, T	W, P, I
3—	J, N, D, Y, F	G, Z
4—	Y, Z, J, Q	C, E?, R?, B?
5—	G, N, A, I, W	Q, L, U

44d. Application of principles of symmetry.—Our next step is to try to determine a few values in each alphabet. In Alphabet 1 we have the following, from the analysis above:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 Cipher—C? I C? M Q V

Let us also set down the values of E already assumed in the remaining alphabets:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1—	C?				I				C?						M		Q		V							
2—					W																					
3—					G																					
4—					C																					
5—					Q																					

FIGURE 15.

We see that by good fortune the letter Q is common to Alphabets 1 and 5, and the letter C is common to Alphabets 1 and 4. If we are dealing with a case in which a mixed component is sliding against the normal component, we can apply the principles of symmetry of position to these alphabets, as outlined on page 53. For example, we may insert the following values in Alphabet 5:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1—	C?				I				C?						M			Q		V						
5—		M		Q		V							C?				I					C?				

The process at once gives three definite values: $\overset{5}{M}_c = B_p$, $\overset{5}{V}_c = G_p$, $\overset{5}{I}_c = R_p$. Let us corroborate them by referring to the frequency table. Since B and G are normally low or medium frequency letters in plain text, we should find that M_c and V_c , their hypothetical equivalents in Alphabet 5, should have low frequencies. As a matter of fact, they do not appear in this alphabet, which corroborates our assumption so far. On the other hand, since $\overset{5}{I}_c = R_p$, if our values derived from symmetry of position are correct, $\overset{5}{I}_c$ should be of high frequency, and it is. The position of C is doubtful; it belongs either under N_p or V_p . If the former is correct, then the frequency of $\overset{5}{C}_c$ should be high, for it would equal N_p ; if the latter is correct, then its frequency should be low, for it would equal V_c . As a matter of fact $\overset{5}{C}_c$ does not occur, and we must conclude that it belongs under V_p . This in turn settles the value of $\overset{1}{C}_c$, for it must now be placed definitely under I_p and removed from beneath A_p .

The definite placement of C now enables us to insert new values in Alphabet 4, and we now have the following:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1—				I				C						M			Q		V						
	2—				W																					
	3—				G																					
	4—	I			C						M			Q		V										
	5—		M		Q		V											I					C			

FIGURE 10.

44e. Substitution of deduced values.—It is high time we inserted our values in the cipher, for by this time it must seem that we have certainly gone too far with work based upon thus far unproven hypotheses. The following results:

MESSAGE.

SIJYU MNVCA ISPJL RBZEY QWYEU LWMGW ICJCI MTZEI MIBKN
 O E E RE E E ER O R O

QWBRI VWYIG BWNBQ QCGQH IWJKA GEGXN IDMRU VEZYG QIGVN
 RE R TE A E E R EN EE E E T R EP

CTGYO BPDBL VCGXG BKZZG IVXCU NTZAO BWFEQ QLFCO MTYZT
 I E T E E E E E R E O

CCBYQ OPDKA GDGIG VPWMR QIIEW ICGXG BLGQQ VBGRS MYJJY
 I E EA T K R E E ENE T E O

QVFWY RWNFL GXNFW MCJKX IDDRU OPJQQ ZRHCN VWDYQ RDGDG
 R E O E NE E TE E E

BXDBN PXFPU YXNFG MPJEL SANCD SEZZG IBEYU KDHCA MBJJF
 O E E E O

KILCJ MFDZT CTJRD MIYZQ ACJRR SBGZN QYAHQ VEDCQ LXNCL
 E O I O E E R E T EE E

LVVCS QWBII IVJRN WNBRI VPJEL TAGDN IRGQP ATYEW CBYZT
 E RE AR E R T E E EN A I

EVGQU VPHYL LRZNQ XINBA IKWJQ RDZYF KWFZL GWFJQ QWJYQ
 EN T E E E E E E RE E

IBWRX
 E

The combinations given are excellent throughout and no inconsistencies appear. We now note the trigraph ^{1 2 3}QWB which is repeated in the following polygraphs above underlined:

1 2 3 4 5 1 . . . 5 1 2 3 4 5 1
 Q W B R I V . . . S Q W B I I I
 R E RT . . . R E A R E

The letter $\overset{3}{B}_0$ is common to both polygraphs, and a little imagination will lead to the assumption of the value $\overset{3}{B}_0 = P_p$, yielding the following:

1 2 3 4 5 1 . . . 5 1 2 3 4 5 1
 Q W B R I V . . . S Q W B I I I
 R E P O R T . . . P R E P A R E

We also note the following polygraph: $\overset{4}{I} \overset{5}{G} \overset{1}{V} \overset{2}{P} \overset{3}{W} \overset{4}{M}$, which looks like the word ATTACK. The frequency tables are consulted to see whether the frequencies given for $\overset{5}{G}_c$ and $\overset{2}{P}_c$ are high enough for T_p and A_p , respectively, and also whether the frequency of $\overset{3}{W}_c$ is good enough for C_p ; we note that they are excellent. Moreover, the digraph $\overset{5}{G}\overset{1}{V}$, which occurs four times, looks like TH, thus making $\overset{1}{B}_c = H_p$. Let us now see whether the insertion of these four new values in our diagram of alphabets brings forth any inconsistencies. The insertion of the value $\overset{2}{P}_c = A_p$ and $\overset{1}{B}_c = H_p$ gives no indications either way, since neither letter has yet been located in any of the other alphabets. The insertion of the value $\overset{5}{G}_c = T_p$ gives us a value common to Alphabets 3 and 5, for we have long ago had the value $\overset{3}{G}_c = E_p$. Unfortunately we find an inconsistency here. The letter I has been placed two letters to the left of G in our mixed component, and has given good results in Alphabets 1 and 5; if the value $\overset{3}{W}_c = C_p$, as obtained above from the assumption of the word ATTACK, is correct, then W, and not I, should be the second letter to the left of G. Which shall we retain? There has been so far nothing to establish the value of $\overset{3}{G}_c = E_p$; we assumed this value from frequency considerations solely. Perhaps it is wrong. It certainly behaves like a vowel, and we may see what happens when we change its value to O_p . The following placements result from our analysis when only two or three new values have been added as a result of the clues afforded by our deductions:

Plain—	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1—			S	I			G	B	C						M	P	Q	R	V	W						
2—	P	Q	R	V	W							S	I	G	B	C									M	
3—	R	V	W							S	I	G	B	C								M		P	Q	
4—	I		G	B	C					M	P	Q	R	V	W										S	
5—	M		P	Q	R	V	W								S	I	G	B	C							

FIGURE 17.

Many new values are produced, and these are inserted throughout the message, yielding the following:

SIJYU MNVCA ISPJL RBZEY QWYEU LWMGW ICJCI MTZEI MIBKN
 CO O BE EMY SR RE EWCH ES ER O R OOP
 QWBRI VWYIG BWNBQ QCGQH IWJKA GEGXN IDMRU VEZYG QIGVN
 REPOR TE AT HE DE RSON EE G O E WO T T ROOP
 CTGYO BPDBL VCGXG BKZZG IVXCU NTZAO BWFEQ QLFCO MTYZT
 I O HA D TSO TH T ED E HE ER E O
 CCBYQ OPDKA GDGIG VPWMR QII EW ICGXG BLGQQ VBGRS MYJJY
 ISP E A G OAT TACKF ROM H ESO TH ONE TROOP O
 QVFWY RWNFL GXNFW MCJXK IDDRU OPJQQ ZRH CN VWDYQ RDGDG
 RD Q SE G H OS E O A NE CE TE ES OT
 BXDBN PXFPU YXNFG MPJEL SANCD SEZZG IBEYU KDHCA MBJJF
 H D Q M T OA C E C T ER E OR
 KILCJ MFDZT CTJRD MIYZQ ACJRR SBGZN QYAHQ VEDCQ LXNCL
 O E O I O OO E S OF CRO R E T EE E
 LVVCS QWBII IVJRN WNBRI VPJEL TAGDN IRGQP ATYEW CBYZT
 DBEP REPAR ED O U POR TA O ECOND H IR
 EVGQU VPYHL LRZNQ XINBA IKWJQ RDZYF KWFZL GWFJQ QWJYQ
 DON TA C E O D E ES E GE ER E
 IBWRX
 ER O

44f. Completing the solution.—Completion of solution is now a very easy matter. The mixed component is finally found to be the following sequence, based upon the word EXHAUSTING:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H
2	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O
3	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q
4	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T
5	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K

FIGURE 18.

Note that the successive equivalents of A_p spell the word APRIL, which is the key for the message. The plain-text message is as follows:

C O Troop B.

Enemy has retired to NEWCHESTER. One troop is reported at HENDERSON MEETING HOUSE; two other troops in orchard at southwest edge of NEWCHESTER. 2d Sq is preparing to attack from the south. One troop of 3d Sq is engaging hostile troop at NEWCHESTER. Rest of 3d Sq is moving to attack NEWCHESTER from the north. Move your Sq into woods east of crossroad 539 and be prepared to support attack of 2d and 3d Sq. Do not advance beyond NEWCHESTER. Messages here.

TREER,
Col.

The preceding case is a good example of the value of the principle of symmetry of position when applied properly to a cryptogram enciphered by the sliding of a mixed component against the normal. We started off with only a very limited number of assumptions and built up many new values as a result of the placement of the few original values in the diagram of the alphabets.

45. Application of foregoing principles—solution of subsequent messages.—Let us suppose that the correspondents are using the same basic or primary component but with different key words for other messages. Can the knowledge of the sequence of letters in the reconstructed primary component be used to solve the subsequent messages?

We found that in the case of a monoalphabet cipher in which a mixed alphabet was used, the process of completing the plain component could be applied to solve subsequent messages in which the same cipher component was used even though the cipher component was set at a different key letter. A modification of the procedure used in that case can be used in this case, where a plurality of cipher alphabets based upon a sliding primary component is used. Let us suppose that the following message passing between the same two correspondents as in the preceding message has been intercepted:

MESSAGE.

SFDZR YRRKX MIWLL AQRLU RQFRT IJQKF XUWBS
 MDJZK MICQC UDPTV TYRNH TRORV BQLTI QBNPR
 RTUHD PTIVE RMGQN LRATQ PLUKR KGRZF JCMGP
 IHSMR GQRFX BCABA OEMTL PCXJM RGQSZ VB

45a. Factoring and conversion into plain component equivalents.—The presence of a repetition of a four-letter polygraph whose

1 2 3 4 5 6 7
 S F D Z R Y R
 R K X M I W L
 L A Q R L U R
 Q F R T I J Q
 K F X U W B S
 M D J Z K M I
 C Q C U D P T
 V T Y R N H T
 R O R V B Q L
 T I Q B N P R
 R T U H D P T
 I V E R M G Q
 N L R A T Q P
 L U K R K G R
 Z F J C M G P
 I H S M R G Q
 R F X B C A B
 A O E M T L P
 C X J M R G Q
 S Z V B

interval is 21 letters suggests a key word of seven letters. There are very few other repetitions, and this is to be expected in a short message with a key of such length.

Let us rewrite the message in groups of seven letters, in columnar fashion, as shown in Figure 19. The letters in each column belong to a single alphabet. Let us convert the letters in each column into their plain component equivalents by setting our reconstructed cipher component against the normal alphabet at any arbitrarily selected point, for example, that shown below:

Plain—ABCDEFGHIJKLMN OPQRSTUVWXYZ
 Cipher—EXHAUSTINGBCDFJKLMOPQRVWYZ

The columns of equivalents are now as follows (fig. 20):

FIGURE 19.

45b. Examination and selection of generatrices.—We found that in the case of a mono-alphabet cipher it was merely necessary to complete the normal alphabet sequence beneath the plain component equivalents and the plain text all reappeared on one generatrix. We found, also, that in the case of a multiple alphabet cipher involving standard alphabets, the plain text equivalents of each alphabet reappeared on the same generatrix, and it was necessary only to combine the proper generatrices in order to produce the plain text of the message. In the case at hand we combine both processes: We complete the normal alphabet sequence beneath the letters of each column and then combine generatrices to produce the plain text. The completion diagrams for the first two columns are as follows (fig. 21):

1 2 3 4 5 6 7
 F N M Z V Y V
 V P B R H X Q
 Q D U V Q E V
 U N V G H O U
 P N B E X K F
 R M O Z P R H
 L U L E M T G
 W G Y V I C G
 V S V W K U Q
 G H U K I T V
 V G E C M T G
 H W A V R J U
 I Q V D G U T
 Q E P V P J V
 Z N O L R J T
 H C F R V J U
 V N B K L D K
 D S A R G Q T
 L B O R V J U
 F Z W

FIGURE 20.

<i>Column 1.</i>	<i>Column 2.</i>
1 FVQUPRLWVGVHIIQZHVDLF	NPDNNMUGSHGWQENCNSBZ
2 GWRVQSMXWHWIIJRAIWEMG	<u>OQEONVHTIIHXRFODOTCA</u>
3 HXSWRTNYXIXJKSBJXFNH	PRFPPOWIUJIYSGPEPUDB
4 IYTXSUÓZYJYKLTCKYGOI	QSGQQPXJVKJZTHQFQVEC
5 JZUYTVPAZKZLMUDLZHPJ	RTHRRQYKWLKAUIRGRWFD
6 KAVZUWQBALAMNVEMAIQK	SUISSRZLXMLBVJSHSXGE
7 LBWAVXRCBMBNOWFNBRL	TVJTTSAMYNMCWKTITYHF
8 MCXBWYSDCNCOPXGOCKSM	UWKUUTBNZONDXLUJUZIG
9 NDYCXZTEDODPQYHPDLTN	VXLVVUCOAPOEYMKVAJH
10 OEZDYAUFEPEQRZIQEMUO	WYMWVDPBQPFZNLWBKI
11 PFAEZBVGFFQFRSAJRFNVP	XZNXXWEQCRQGAOXMCLJ
12 QGBFACWHGRGSTBKSGOWQ	YAOYXXFRDSRHPYNYDMK
13 RHCGBDXIHSHTUCLTHPXK	ZBPZZYGSETSICQZOZENL
14 SIDHCEYJITIUVDMUIQYS	ACQAAZHTFUTJDRAPAFOM
15 TJEIDFZKJUJVVENVJRZT	BDRBBAIUGVUKESBQBGPN
16 UKFJEGALKVKWXFOWKSAU	CESCCBJVHWVLFTRCHQO
17 VLGKFHBMWLXYGPXLTBV	DFTDDCKWIXWMGUDSDIRP
18 WMHLGICNMXYZHQMUCW	EGUEEDLXJYXNHVETEJSQ
19 XNIMHJDONYNZAIRZNVDX	FHVFFEMYKZYOIWFUFKTR
20 YOJNIKEPOZOABJSAOWEY	GIWGGFNZLAZPJXGVGLUS
21 ZPKOJLFPAPBCKTBPXFZ	HJXHHGOAMBAQKYHWHMVT
22 AQLPKMGRQBQCDLUCQYGA	IKYIIHPNBCBRLZIXINWU
23 BRMQLNHSRCRDEMVDZHB	JLZJJIQCDCSMAJYJOXV
24 <u>CSNRMOITSDSEFNWESAIC</u>	KMAKKJRPEDTNBKZKPYW
25 DTOSNPJUTETFGOXFTBJD	LNBLKSEQFEUOCLALQZX
26 EUPTOQKVUFUGHPYGUCKE	MOCMMLTFRGFVPDMBMRAY

FIGURE 21.

45c. **Combining the selected generatrices.**—After some experimenting with these generatrices we put together the 24th generatrix of Column 1 and the 2d of Column 2, which yields the digraphs shown in Figure 22. The generatrices of the subsequent columns are examined to select those which may be added to these already selected in order to build up the plain text. The results are shown in Figure 23.

1 2
C O
S Q
N E
R O
M O
M O
N O
N I
V T
H T
S T
D I
S H
E X
F R
N F
W O
E D
S O
A T
I C
C A

1 2 2 4 5 6 7
C O F I R S T
S Q R A D R O
N E N E M Y T
R O O P D I S
M O U N T E D
O N H I L L F
I V E N I N E
T H R E E W E
S T O F G O O
D I N T E N T
S H X L I N E
E X T E N D S
F R O M C O R
N F I E L D T
W O H U N D R
E D Y A R D S
S O U T H X I
A T T A C K R
I C H A R D S
C A P T

This process is a very valuable aid in the solution of messages after the primary component has been recovered as a result of the longer and more detailed analysis of the frequency tables of the first message intercepted. Very often a short message can be solved in no other way than the one shown, when the primary alphabet is completely known.

We may be interested to find the key word for the message. All that is necessary is to set the mixed component of the cipher alphabet underneath the plain component so as to produce the cipher letter indicated as the equivalent of any given plain-text letter in each of the alphabets. For example, in the first alphabet we note that $C_p = S_c$. Setting the two components under each other so as to bring S of the cipher component beneath C of the plain component, thus:

Plain.

ABCDEF GHIJKLMNOPQRSTUVWXYZ ABCDEF GHIJKLMNOPQRSTUVWXYZ

EXHAUSTINGBCDFJKLMOPQ RVWYZ.

Cipher.

FIGURE 22.

FIGURE 23.

we note that $A_p = A_c$. Hence, the first letter of the key word to the message is A. The 2d, 3d, 4th, ... 7th key letters are found in exactly the same manner, and we have the following:

When C O F I R S T equals
S F D Z R Y R then A_p successively equals
A Z I M U T H

46. Application of foregoing principles—solution by reconstruction of key.—Occasionally one may encounter a cryptogram which is so short that it contains no recurrences of even digraphs, and thus gives no indications of the number of alphabets involved. If the sliding mixed component is known one may apply the method illustrated on page 50, assuming the presence of a likely word, and checking it against the text and the sliding components to establish a key, if the correspondents are using key words. For example, suppose that we assume the presence of the word ENEMY in the message directly above and proceed to check it against an unknown key word, using the already reconstructed mixed component sliding against the normal and starting with the first letter of the cryptogram in this manner:

If SFDZR equals ENEMY, then the successive equivalents of A, equal XENFW.

The sequence XENFW spells no intelligible word. We therefore shift the location of our assumed word ENEMY one letter forward in the cipher text, and try again, just as was explained on page 50. When we try the group AQRLU we obtain as the key letters ZIMUT, which suggests the word AZIMUTH. The method must yield solution when a correct word is assumed and correctly placed.

So far we have considered only the case in which the cipher alphabets employed in a multiple alphabet cipher are secondary alphabets produced by the sliding of a mixed component against the normal sequence. The next step in complexity concerns the case in which the cipher alphabets are secondary alphabets produced by the sliding of two mixed components against each other.

SECTION XI.

SOLUTION OF PERIODIC POLYALPHABET CIPHERS USING MIXED ALPHABETS.

CASE II.—Both components are mixed sequences.

	Paragraph.
Identical mixed components.....	47
Use of a quadricular table instead of sliding components.....	47a
Examination of secondary alphabets.....	47b
Reconstruction of the primary component.....	47c
Method of solution after primary component has been reconstructed.....	47d
Reversed identical mixed components.....	47e
Nonidentical mixed components.....	48

47. Identical mixed components.—This case may be further subdivided into two cases: (1) The mixed sequences are identical in both components, and (2) the mixed sequences are different. The first case will now be considered.

It is often the case that the mixed sequences are derived from an easily remembered word or phrase, so that the sliding strips can be reproduced at any time from memory. Thus, for example, given the key word QUESTIONABLY, the following mixed sequence is derived:

QUESTIONABLYCDFGHJKMPRVWXZ

By sliding this sequence against itself, a series of 26 secondary mixed alphabets may be produced. For example, by setting the two sliding strips against each other in the two positions shown below, the cipher alphabets labeled (1) and (2) given by the two settings are seen to be different.

Plain component.

QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ

QUESTIONABLYCDFGHJKMPRVWXZ

Cipher component.

Secondary alphabet.

(1) Plain—ABCDEFGHIJKLMN**OP**QRSTUVWXYZ
Cipher—CDHJOKMPBRVFWYLXTZNAIUEGS

Plain component.

QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ

QUESTIONABLYCDFGHJKMPRVWXZ

Cipher component.

Secondary alphabet.

(2) Plain—ABCDEFGHIJKLMN**OP**QRSTUVWXYZ
Cipher—HJPRLVWXDZQKUGFEASYCBTIOMN

In enciphering a message by such sliding strips, a key word is used to designate the particular positions in which the strips are to be set, the same as was the case in previous examples of the use of sliding components. The method of designating the positions is, however, slightly different, reasons for which will appear in the succeeding paragraph. In the methods heretofore given, the key letter, as located on the cipher component, was set opposite A, as located on the plain component; in other words, if T was the key letter, then the two sliding strips were set so that $A_p = T_c$. In this case, however, where identical mixed sliding components are used, the key letter is set opposite the *first* letter of the sequence upon which the primary components are based; that is, if T is the key letter, then the sliding strips are set so that $Q_p = T_c$, in the case of the mixed components shown above. Hence, in the first of the two examples above, the key letter being T, then T_c is set opposite Q_p ; in the second of these examples, the key letter being A, then A_c is set opposite Q_p .

47a. Use of a quadricular table instead of sliding components.—Very frequently a quadricular or square table is employed by the correspondents, instead of the ordinary sliding strips, but the results are the same. The square table based upon the word QUESTIONABLY is shown in Table 4. It will be noted that the table does nothing more than set forth the successive positions of the two primary sliding components, and the top line of the table is the plain component, the successive horizontal lines below it, the cipher components of the various alphabets formed. The usual method of employing such a table is to take as the cipher equivalent of a plain-text letter that letter which lies at the intersection of the vertical column headed by the plain-text letter and the horizontal row begun by the key letter. For example, the cipher equivalent of E_p with key letter T is the letter O_c or $E_p (T_k) = O_c$. The method given in paragraph 47, for determining the cipher equivalents by means of the two sliding strips yields the same results as does the square table.

TABLE 4.

Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z
U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q
E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U
S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E
T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S
I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T
O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I
N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N
B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A
L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B
Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L
C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y
D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C
F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D
G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G
J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H
K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J
M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K
P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M
R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P
V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V
X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W
Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X

47b. Examination of secondary alphabets.—Note that the two secondary alphabets given in paragraph 47 show no *external* resemblance or similarity as a result of their having been produced from the same primary components. It is therefore impossible in this

case to apply the same principles of symmetry of position to the various cipher alphabets being analyzed in solving a cryptogram, as were applied in the preceding case. Nevertheless modified principles of symmetry of position can be applied, but they are somewhat complex, and do not fall within the scope of this book. It is only in advanced work, and then only in special cases, where the possibility of the application of modified principles of symmetry of position is of importance in the analysis of complex cryptograms. Symmetry of position which, though present, is not manifested externally is termed *indirect symmetry*.

The beginner will find it necessary to solve the individual alphabets without recourse to any principles of symmetry of position in those cases where the cipher alphabets are secondary alphabets produced by mixed components, or by a table such as that illustrated in Table 4. The analysis of such alphabets is exactly the same as before, viz, by considerations of frequency and repetitions of digraphs, trigraphs, and polygraphs.

47c. Reconstruction of the primary component.—There is, however, one principle which is useful and easy to demonstrate. After a cryptogram has been solved by a detailed analysis, and there is at hand one complete or nearly complete secondary alphabet, it often becomes desirable to find the primary components, so that subsequent messages can be solved more easily. Suppose, for example, that the analysis of a message has yielded the following cipher alphabet:

Plain—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
 Cipher—J K R V Y W X Z F Q U M E H G S B T C D L I O N P A

We suspect that this is a secondary alphabet produced by two primary identical mixed components and wish to reconstruct the latter. Construct a chain of alternate plain text=cipher=plain text values, beginning at any point and continuing until the chain has been completed. Thus, for example, beginning with $A_p=J_c$, $J_p=Q_c$, $Q_p=B_c$, and dropping out the letters common to successive pairs, we have the sequence AJQB By completing the chain we establish the following sequence of letters:

AJQBKULMEYPSCRTDVIFWOGXNHZ

This chain consists of 26 letters, and when slid against itself will produce exactly the same secondary alphabets as do the primary components based upon the word QUESTIONABLY. To demonstrate that this is the case, compare the secondary alphabets given by the two settings of the externally different components shown on page 73.

Plain component.
 QUESTIONABLYCDFGHJKMPRVWXXZQUESTIONABLYCDFGHJKMPRVWXXZ
 QUESTIONABLYCDFGHJKMPRVWXXZ
 Cipher component.

Secondary alphabet.
 (1) Plain—ABCDEFGHIJKLMNQRSTUWXYZ
 Cipher—JKRVYWXZFQUMEHGSBTDLIONPA

Plain component.
 AJQBKULMEYPSCRTDVIFWOGXNHZAJQBKULMEYPSCRTDVIFWOGXNHZ
 AJQBKULMEYPSCRTDVIFWOGXNHZ
 Cipher component.

Secondary alphabet.
 (2) Plain—ABCDEFGHIJKLMNQRSTUWXYZ
 Cipher—JKRVYWXZFQUMEHGSBTDLIONPA

Since the sequence AJQBK . . . gives exactly the same equivalents in the secondary alphabets as the sequence QUEST . . . gives, it is termed an *equivalent primary component*. The real or original primary component can be obtained from the equivalent primary component by a rearrangement of the letters of the latter according to a very simple system. Find three letters in the sequence such as are likely to have formed an unbroken sequence in the original primary component, and see if the interval between the first and second is the same as that between the second and third. Such a case is presented by the letters W, X, and Z in the equivalent primary component above; the distance or interval between them is two letters. Continuing the chain by adding letters two intervals removed we have the following:

WXZQUESTIONABLYCDFGHJKMPRV

This corresponds to the original primary component.

Not all of the 26 secondary alphabets of the series yielded by two sliding primary components may be used to develop a complete equivalent primary component. If examination be made, it will be found that only 13 of these secondary alphabets will yield complete equivalent primary components. For example, the following secondary alphabet, which is also derived from the primary components based upon the word QUESTIONABLY will not yield a complete chain of 26 plain text=cipher=plain text equivalents:

Plain—ABCDEFGHIJKLMNQRSTUWXYZ
 Cipher—CDHJOKMPBRVFWYLXTZNAIQUEGS

Equivalent primary component:

ACHPXEOLFKVQTACH . . .

We see that only 13 letters of the chain have been established before the sequence begins to repeat itself. It is evident that we

have here exactly one-half of the chain. We may build up the other half by beginning with a letter not in the first half. Thus:

BDJRZSNYGMWUIBDJ...

It is now necessary to distribute the letters of each half-sequence within 26 spaces, to correspond with their placements in a complete alphabet. This can only be done by allowing between the letters of one of the half-sequences a constant *odd* number of spaces. Distributions are therefore made upon the basis of 3, 5, 7, 9, ... spaces. Select that distribution which most nearly coincides with the distribution to be expected in a key-word component. Thus, for example, with the first half-sequence the distribution selected is the one made by leaving three spaces between the letters; it is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	.	L		C		F		H		K		P		V		X		Q		E		T		O	

Now interpolate, by the same constant interval (three in this case), the letters of the other half-sequence. Noting that the group F-H appears in the foregoing distribution, it is apparent that G of the second half-sequence should be inserted between F and H. The letter which immediately follows G in the second half-sequence, viz; M, is next inserted in the position three spaces to the right of G, and so on, until the interpolation has been completed. This yields the original primary component, which is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N

47d. Method of solution after primary component has been reconstructed.—After the primary components have been reconstructed it is possible to solve all cryptograms enciphered by them by applying the process of completing the plain component sequence in the manner illustrated in previous cases similar to this. It is not necessary, however, to convert the cipher letters into their plain component equivalents because the completion sequence that must be employed is the primary sequence itself; in other words the plain component coincides with the primary component. An example will serve to illustrate the method. Suppose we know that the correspondents are using the primary components reconstructed above, viz: QUESTIONABLYCDFGHJKMPRVWXZ, and we have intercepted the following message:

PCEGC GTHIE EPSGK FSYQU FHDOL VSNXF
GDJLZ NFGCQ WAFYQ XKIGF RTCQV KHLJY
 FUAM

The two repetitions FG and CQ factor for 7 alphabets.
The message is rewritten in groups of seven letters, in columns.

1 2 3 4 5 6 7
P C E G C G T
H I E E P S G
K F S Y Q U F
H D O L V S N
X F G D J L Z
N F G C Q W A
F Y Q X K I G
F R T C Q V K
H L J Y F U A
M

The completion sequence, using the primary component already reconstructed from a previous message, is applied to each column, and then the generatrices are assembled to yield intelligible text. The completion diagrams for the first three columns are as follows:

1	PHKHXNFFHM	CIFDFFYRL	EESOGGQTJ
2	RJM JZAGGJP	DOGFGGCVY	SSTNHHUIK
3	VKPKQBHHKR	FNHGHHDWC	TTIAJJEOM
4	WMRMULJJMV	GAJHJJFXD	IIOBKKSNP
5	XPVPEYKKPW	HBKJKKGZF	<u>OONLMMTAR</u>
6	ZRWRSCMMRX	JLMKMMHQG	NNAYPPIBV
7	QVXVTDPPVZ	KYPMP PJUH	AABCRRRLW
8	UWZWI FRRWQ	MCRPRRKEJ	BBLDVVNYX
9	EXQXOGVVXU	PDVRVVM SK	LLYFWWACZ
10	SZUZNHWWZE	RFWVWWPTM	YYCGXXBDQ
11	TQEQA JXXQS	VGXWXXRIP	CCDHZZLFU
12	IUSUBKZZUT	WHZXZZVOR	DDFJQQYGE
13	OETELMQQEI	XJQZQQWNV	FFGKUUCHS
14	NSISYPUUSO	ZKUQUUXAW	GGHMEEDJT
15	<u>ATOTCREETN</u>	QMEUEEZBX	HHJPSSFKI
16	BINIDVSSIA	UPSESSLZ	JJKRTTGMO
17	LOAOFWTTOB	ERTSTTUYQ	KKMVIHPN
18	YNBNGXIINL	SVITIIIECU	MMPWOOJRA
19	CALAHZOOAY	<u>TWOIOOSDE</u>	PPRXNNKVB
20	DBYBJQNNBC	IXNONNTFS	RRVZAAMWL
21	FLCLKUAALD	OZANAAIGT	VVWQBBPXY
22	GYDYMEBBYF	NQBABBBOHI	WWXULLRZC
23	HCFCPSLLCG	AULBLLNJO	XXZEYYVQD
24	JDGDRTYYDH	BEYLYYAKN	ZZQSCCWUF
25	KFH FVICCFJ	LSCYCCBMA	QQUTDDXEG
26	MGJGWODDGK	YTDCDDL PB	UUEIFFZSH

FIGURE 24.

When we select and combine the 15th, 19th, and 5th generatrices of these three diagrams, we obtain the trigraphs shown in Figure 25. Completion of the message follows easily enough. The key letters can be obtained by noting the equivalent of Q_p , which is the *first* letter of the primary components, in each of the alphabets. The first three key letters are as follows:

Alphabet 1— $Q_p = C_c$
 Alphabet 2— $Q_p = A_c$
 Alphabet 3— $Q_p = V_c$

1 2 3
 A T O
 T W O
 O O N
 T I L
 C O M
 R O M
 E S T
 E D A
 T E R
 N

FIGURE 25

The syllable CAV suggests the word CAVALRY, and this word applied to the message completely solves it.

47e. Reversed identical mixed components.—The identical components in the preceding discussion progress in the same direction. They may, however, progress in opposite directions, in which case a series of 26 reciprocal secondary alphabets is produced and is evidenced by the appearance of reciprocal values in each of the cipher alphabets employed in the encipherment of a message. The suspicion or knowledge that the cipher alphabets are reciprocal alphabets is, of course, a great aid in solution. The reconstruction of the primary components is somewhat different in this case and requires two secondary alphabets, although in certain cases it can be done with only one, after a more or less lengthy experimentation, providing the primary component is actually a key-word alphabet. The use of reversed identical components is so infrequent, however, that a discussion of the methods to be applied in reconstructing the original or an equivalent primary component is not considered to be of sufficient importance to warrant its inclusion in this book. It may be added, however, that after the reconstruction has been effected the solution of subsequent messages can be achieved by a recourse to the principle of completing the plain component sequence.

48. Nonidentical mixed components.—A further complexity in polyalphabet ciphers employing sliding components is introduced when the two components are different mixed sequences. Where considerable text is available, solution may be achieved by ordinary frequency principles, paying no attention to the fact that the secondary alphabets are interrelated. Where only a small amount of text is at hand, a rather complicated process of reconstructing the primary components by recourse to the principles of indirect symmetry of position, a process which is too complicated to explain here, can be applied. The analysis of the cipher text and the reconstruction of the primary components progress simultaneously, one aiding the other.

It is then desirable to reconstruct the primary components completely, even though the analysis might have been done by reference to the simple principles of frequency, for the solution of subsequent messages now becomes a simple matter. This reconstruction may be effected by a modification of the principles illustrated on page 72 in connection with identical mixed components. Two secondary alphabets are necessary for the process, two separate chains or series of equivalents being established, and the two components assembled by reference to the common letters of pairs that are to be joined.

After the primary components have been reconstructed the solution of subsequent messages in different keys may be attained by the method of completing the plain component sequence. The cipher letters must first be converted into their plain component equivalents and then the plain component (mixed) sequence is to be completed beneath the converted letters. Selection and assembling of generatrices is the same as usual.

SECTION XII.

SOLUTION OF PROGRESSIVE POLYALPHABET CIPHERS.

	Paragraph.
General remarks.....	49
Solution of an example using standard alphabets.....	50
The case of mixed alphabets.....	51

49. General remarks.—We shall now take up very briefly the other type of periodic polyalphabet system, viz, that previously mentioned under the name progressive alphabet system, in which the members of a whole series of cipher alphabets are employed one after the other in a definite sequence. These are usually encountered in some types of machine ciphers and are often very difficult to solve unless one is in possession of a duplicate machine or of the alphabets used. The length of the period is usually equal to the total number of different secondary alphabets employed in the system, and nearly always the sequence in which the alphabets are employed varies in different messages, this constituting one of the principles of the secrecy of the system. It is necessary, therefore, to intercept either one very long message or a series of messages in the same key in order to have sufficient material to form a basis for analysis.

50. Solution of an example using standard alphabets.—There is one variety of such a system that can be solved most readily. It is the one in which standard alphabets are used, and one of the sliding components is advanced one letter regularly throughout the message. For example, if the message starts off with $A_p = K_c$, the first letter is enciphered by that alphabet; the next letter is then enciphered by the next alphabet, in which $A_p = L_c$; the third letter, by the alphabet in which $A_p = M_c$, and so on. The solution is exceed-

ingly simple. We found that all monoalphabet ciphers in which standard alphabets are employed can be solved by completing the normal alphabet sequence beneath the cipher letters, and that the plain text all reappears on the same generatrix. If the alphabets are shifted regularly one space at a time, then the plain text will reappear on a *diagonal* line instead of a horizontal line. If, in setting up the letters on the sliding strips when applying the completion sequence process to the cipher letters, we set up the cipher letters on a diagonal line, each letter one line above or below the preceding letter, then the plain text will reappear on the same generatrix instead of on a diagonal line. The example which follows is a cipher which was submitted as "indecipherable" and it is only one of the many instances in which the claims of indecipherability are found to be wholly unwarranted after only a very superficial analysis.

MESSAGE.

THXYT GHGRD MGAGB UPQOQ PZVYQ XYVED
 VJLEE RZWFG NLHUQ RNFEN XPGIH UFEFN
 MINKW UONTS QQPSZ PKWJE PFRRO EAXHO
 FUXNX UQWGF DQZTD RCIKA GZNMV WZNAM
 JGFCQ LZZCB ZMBBK URIPK

Trial of direct standard alphabets, using sliding strips, yields no results. We therefore convert the first two or three groups into their direct standard alphabet equivalents and then set up these letters on a diagonal line running downwards. Notice the plain text reappearing on one generatrix, Figure 26, beginning GRAY COMMANDING.

Cipher letters—T H X Y T G H G R D M G A G
 Direct alphabet equivalents—A M W V A N M N C Q H N T N

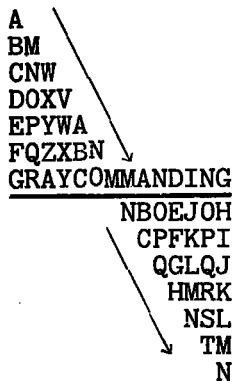


FIGURE 26.

The only further difficulty in deciphering this message is encountered in finding the meaning of certain abbreviations such as "KG," "XX," "QQ," etc., which stand for punctuation, and for numbers.

51. The case of mixed alphabets.—It is obvious that the only reason why such a system is so simple to solve is that the cipher alphabets are known alphabets. If two mixed components were used to encipher the message, solution of a single short message would be an extremely difficult matter, if at all possible.

The student might therefore conclude that such a system is practically indecipherable, but such is far from the case. If a sufficient amount of text is available (and the use of the term "system" presupposes that it would be used for regular traffic, thus making many messages in the same alphabets available for interception), the messages can be reduced, though only after much study. The method can not be described here in detail, but suffice it to say that the messages would be superimposed (see Par. 56a.) in such a manner that the letters thus brought into each column would belong to the same secondary alphabet, and analysis of the individual alphabets by frequency methods, together with a reference to principles employed in the reconstruction of primary components, would yield solution.

Before leaving the subject of periodic systems a few words are to be said about a minor variation of method. We have noted in all the examples given that the cipher alphabets change with successive single letters of the text, but it is of course possible to change the alphabets with successive pairs of letters, or successive groups of definite length, say five to ten or more letters. This variation in method does not suppress cyclic phenomena in the cryptograms, and in certain cases may even increase the manifestations of periodicity.

SECTION XIII.

SOLUTION OF NONPERIODIC POLYALPHABET CIPHERS.

	Paragraph.
Difficulties of solution.....	52
Solution of an example using direct standard alphabets.....	53
The case of reversed standard alphabets.....	54
The case of mixed alphabets.....	55
Ciphers employing a nonrepeating key.....	56
Solution by superimposition of messages.....	56a
Solution by simultaneous decipherment and reconstruction of key.....	56b
Solution of special cases.....	56c

52. Difficulties of solution.—We have seen that the existence of cyclic or periodic phenomena in cryptograms is of great aid in solution, because they afford clues to the enemy cryptanalyst as to how he can regroup, distribute, or assemble the letters of a cryptogram in order to bring together letters which have been enciphered by identical cipher alphabets, and thus provide sufficient material upon which to base an analysis by frequency. It would therefore seem obvious that the elimination of cyclic or periodic phenomena in the cryptograms should remove these clues to solution, and in general this is true.

There are, indeed, more methods and systems of encipherment which greatly suppress or entirely eliminate any periodicity in the resultant cryptograms than systems which do produce manifestations of periodicity, and most of the very complex types of ciphers fall into this class of nonperiodic systems. It will be impossible to go into detail with illustrations and methods employed in the encipherment and analysis of these cryptograms, for a complete exposition of only two or three of them would alone fill a book of this size, and there are dozens of methods. Many of them are written by means of cipher machines and apparatus of more or less complicated nature. In every case, however, a knowledge of the cipher alphabets involved in the system constitutes the greatest aid in solution. When, therefore, standard alphabets are employed solution may be exceedingly simple. Let us consider a case.

53. Solution of an example using direct standard alphabets.—Two correspondents decide to use the Vigenère Table with a key word or a key phrase, but instead of changing the alphabet with successive letters they agree to change it with successive complete words. Since word lengths are very irregular, even if repetitions do occur, the intervals between these repetitions will show no constant factor, and, therefore, it will be impossible to determine the length of the key. This reasoning is strictly correct, but unfortunately for the correspondents, the length of the key or the key itself is in this case of no great interest to the cryptanalyst. The cryptogram can be solved by using sliding strips bearing the normal alphabet. The successive words of the plain text will reappear on different generatrices and it will only be necessary to find these words and assemble them properly. If reversed standard alphabets have been employed, the cipher letters must first be converted into their direct alphabet equivalents. An example follows.

MESSAGE.

QJUCC BERNP UVATG VSWWJ GSVK

As usual, the first test we apply is that of using our normal alphabet strips to complete the normal alphabet sequence. At once the message reappears, the plain-text words being found on various generatrices, and it being merely a matter of inspection to determine them.

MESSAGE.

QJUCC	BERNP	UVATG	VSWWJ	GSVK
RKVDD	CFSOQ	VWBUH	WTXXK	HTWL
SLWEE	DGTPR	WXCVI	XUYYL	IUXM
TMXFF	EHUQS	XYDWJ	YVZZM	JVYN
UNYGG	FIVRT	YZEXK	ZWAAN	KWZO
VOZHH	GJWSU	ZAFYL	AXBBO	LXAP
WPAII	HKXTV	ABGZM	BYCCP	MYBQ
XQBJJ	ILYUW	BCHAN	CZDDQ	NZCR
YRCKK	JMZVX	CDIBO	DAEER	OADS
ZSDLL	KNAWY	DEJCP	EBFF	A_p-S_c
ATEMM	LOBXZ	EFKDQ	FCCG	
BUFNN	MPCYA	FGLER	GDHH	
CVGOO	NQDZB	GHMFS	HEII	
DWHP	A_p-O_c REAC	HINGT	IFJJ	
EXIQ		A_p-N_c U	JGKK	
FYJR		V	KHLL	
GZKS		W	LIMM	
HALT		X	MJNN	
A_p-J_c		Y	NKOO	
		Z	OLPP	
		A	PMQQ	
		B	QNRR	
		C	ROSS	
			A_p-E_c	

FIGURE 27.

If we find the key letters for the successive words, using A_p as the determining letter, we have the word JONES. A long key could be used and the solution would be just as simple.

54. The case of reversed standard alphabets.—Had no good results been obtained by this test, the next step would have been to convert the cipher letters into their reversed standard alphabet equivalents before applying the completion process to the columns. It is of course possible to encipher groups of definite length by the same key letter, instead of word lengths, but the method of solution would be exactly the same. No great difficulty would be experienced in picking up the plain text on the various generatrices.

55. The case of mixed alphabets.—The case is very much more complicated, however, if the correspondents employ mixed alphabets which are unknown to the enemy cryptanalyst. Whereas in the preceding case even a very short message can be solved without difficulty because the cipher alphabets are known sequences, in this

case a large amount of text would be necessary for solution, because the cipher alphabets are unknown sequences. The method of solution would require a careful search for repetitions, and, of course, repetitions of complete words would be inevitable; assumptions of plain-text values for these repetitions would be made; and then an attempt to reconstruct the primary components upon which the cipher alphabets are based would be necessary. If a number of different cipher equivalents for frequently repeated words such as AND, THE, FOR, etc., can be definitely determined from repetitions in the cryptogram, reconstruction of the primary components is fairly easy by recourse to the principles of direct symmetry of position, if only one of the components is a mixed sequence, and to the more complex principles of indirect symmetry, if both components are mixed sequences.

56. Ciphers employing a nonrepeating key.—One of the most common nonperiodic systems is that involving the use of a *non-repeating key*, sometimes called a *continuous* or a *running key*, to encipher the message. The key, instead of repeating itself over and over again, as in the multiple alphabet system, does not repeat itself at all; it often consists of the plain text of a book agreed upon by the correspondents, the exact starting point of the key being predetermined, or indicated in some hidden manner in the cryptogram itself. Often each plain-text letter, or in some cases, each cipher letter constitutes the key letter for enciphering the succeeding plain-text letter, thus eliminating the necessity of using the text of a book for the nonrepeating key, and this method is somewhat more easy to reduce than the one in which a book is used. In all cases the cipher alphabets may be either standard or mixed alphabets.

56a. Solution by superimposition of messages.—The analysis of cryptograms enciphered by a nonrepeating key is extremely interesting but often exceedingly difficult. However, the difficulty becomes very much reduced if many messages are available for study. In fact, in certain cases the solution of a series of messages enciphered by the same nonrepeating key is exceedingly simple even though a single message when considered alone may be extremely difficult to solve. The method consists in searching for repetitions or identities of digraphs, trigraphs, and polygraphs in the several messages, then superimposing the messages so as to bring these repetitions in the same columns, and solving the columns as in ordinary polyalphabet cryptograms. If standard alphabets are used, the analysis of the columns becomes a very easy matter by applying the process of completing the normal alphabet sequence, as explained on pages 47-49. An example will serve to make the process clear. Suppose the following messages have been intercepted on the same day:

1. VUFWZ LLNYZ YAGBS LAWBH PHROM CVJWA NMYNP
 TLJMD GDUFR AXUNT
2. HLZQA VSZVJ MLWFT POABE MLJQH PDPTL JMDGE
 UGSUM KLIDF NWMQA XTYYA EAAQY IWRN
3. BWQZV NNLWF TPZXG NIDJP NHECQ RPHNC SF
4. LMWRW HYMQN LPQAY NCITW ANMYA QEIKS NAAUG
 DQW

Note the underlined identities found in the four messages, by reference to which the messages may be superimposed in the following manner:

	1	5	10	15	20	25	30	35	40	45	50																																										
1.	V	U	F	W	Z	L	L	N	Z	Y	A	G	B	S	L	A	W	B	H	P	H	R	O	M	C	V	J	W	A	<u>N</u>	<u>M</u>	<u>Y</u>	<u>N</u>	<u>P</u>	<u>T</u>	<u>L</u>	<u>J</u>	<u>M</u>	<u>D</u>	<u>G</u>	<u>D</u>	<u>U</u>	<u>F</u>	<u>R</u>	<u>A</u>	<u>X</u>	<u>U</u>	<u>N</u>	<u>T</u>				
2.						H	L	Z	Q	A	V	S	Z	V	J	<u>M</u>	<u>L</u>	<u>W</u>	<u>F</u>	<u>T</u>	P	O	A	B	E	M	L	J	Q	H	P	D	P	T	L	J	M	D	G	E	U	G	S	U	M	K	L	I					
3.																B	W	Q	Z	V	N	N	L	W	F	T	P	Z	X	G	N	I	D	J	P	N	H	E	C	Q	R	P	H	N	C	S	F						
4.																L	M	W	R	W	H	Y	M	Q	N	L	P	Q	A	Y	N	C	I	T	W	A	N	M	Y	A	Q	E	I	K	S	N	A	A	U	G	D	Q	W

FIGURE 28.

The *columns* of letters found by this superimposition of messages now constitute the single cipher alphabets of a polyalphabet substitution cipher. For example, the letters of the twelfth column, AABW, are now in the same alphabet; the letters of the thirteenth column, GVWR, are in another alphabet; and so on. If it is suspected that standard alphabets have been used, we may apply the process of completing the normal alphabet sequence to the letters of each column, select the best generatrix of each column, and assemble the selected generatrices to make intelligible text. If unknown mixed alphabets are used, then the letters of each alphabet must be solved on the basis of frequency, and, of course, a considerable body of text is necessary for solution.

The student should take careful note of the principle of superimposing a series of messages in the same key, for it is the principle which underlies the solution of many complex systems of cipher and is especially important in the analysis of systems used in military cryptography, because here we have the condition where many messages in the same key are easily intercepted by the enemy.

56b. Solution by simultaneous decipherment and reconstruction of key.—It is to be noted in passing that sometimes a single message enciphered by a key which, though it is nonrepeating, constitutes plain text itself, may be solved by reconstructing the key and deciphering simultaneously, using one as a check upon the other. This, of course, can be done only when the cipher alphabets or the primary components that have been employed are known. For example, in the case of the four messages above, if we consider

only the first message, and assume that the word PLATOON occurs somewhere in the message, and that reversed standard alphabets have been employed, we proceed to test out each group of seven letters in the manner explained on page 50. Thus:

If VUFWZLL equals
PLATOON, then A_p equals
KFFPNZY on the basis of reversed standard alphabets.

The sequence KFFPNZY is, of course, unintelligible, so we move our assumed word one letter to the right, and proceed as before. When we come to the group PTLJMDG we find the following:

If PTLJMDG equals
PLATOON, then A_p equals
EELCART

Here we have intelligible text, suggesting the word REELCART. Applying the key letter $A_p = R_c$ (given by R of REEL) to the letter N_c immediately preceding the group PTLJMDG, we obtain E_p as the plain-text letter immediately preceding PLATOON. This in turn suggests the word ONE to precede the word PLATOON, and this gives TAL for the key, suggesting the word REGIMENTAL, and so on. Thus, the decipherment and the reconstruction of the key proceed simultaneously, one suggesting assumptions for the other, until the whole message has been reduced. It is admittedly a laborious process, but will produce solution in time.

56c. Solution of special cases.—In those systems wherein each plain-text letter becomes the key letter for enciphering the succeeding plain-text letter, assumptions of likely words in the plain text enable a reconstruction to be effected. In fact, no assumption of words is really necessary if one can get a start from the very beginning of the message. Consider a message beginning with the letters STWZL XHZRX, etc., enciphered by means of reversed standard alphabets.

If we convert these letters into their direct alphabet equivalents, and then set a series of sliding strips bearing the normal alphabet sequence in such a manner that A of each strip is opposite the converted cipher letters, the plain text will immediately reappear on one generatrix. Thus:

Cipher letters—STWZLXHZRX
Direct alphabet equivalents—IHEBPD TBJD

Setting the sliding strips so that A on the first strip equals I on the second strip; A on the second equals H on the third; A on the third equals E on the fourth, and so on, we have what is shown in Figure 29.

A-P Note the word HOSTILE appearing on one genera-
 A-B Q trix. Solution is thus exceedingly rapid. There is
 B C R absolutely no security in a system of this type, as
 C D S may readily be seen. Even the employment of
 D E T mixed alphabets would not greatly complicate so-
 A-E F U lution, for reflection will show that the letters which
 B F G V immediately follow identical plain-text letters must
 C G H W be in the same alphabet and thus repetitions of
 D H I X cipher letters will be produced in the cryptogram.
 E I J Y These repetitions and the frequency tables can be
 F J K Z used as a basis for solution.
 G K L A Where each cipher letter becomes the key for
 A-H L M B enciphering the succeeding plain-text letter the case
 B I M N C is still more simple to solve, for it is apparent that
 C J N O D the letters which immediately follow identical cipher
 D K O P E letters belong to the same cipher alphabet, and, where
 E L P Q F a sufficient amount of text is available, mixed
 F M Q R G alphabet ciphers of this nature may be solved upon
 G N R S H the basis of frequency alone.
 H O S T I Besides being solvable, all such systems are ex-
 A-I P T U J ceedingly slow as regards encipherment and de-
 B J Q U V K cipherment, and give rise to messages which often
 C K R V W L can not be deciphered promptly on account of the
 many errors which creep in. They are not at all suit-
 able for military cryptography.
 Z Z Z Z Z Z

FIGURE 29.

SECTION XIV.

MISCELLANEOUS SUBSTITUTION METHODS.

	Paragraph.
Number ciphers.....	57
Digraphic substitution.....	58
Playfair cipher.....	58a
Trigraphic and polygraphic substitution.....	59
Machine ciphers.....	60
Concluding remarks on polyalphabet substitution ciphers.....	61

57. Number ciphers.—The various systems discussed in the preceding pages by no means exhaust the different methods that may be used in enciphering a message on the substitution principle, and each one that was presented may be modified in minor details indefinitely. However, under this subdivision we may say a few words about the most common variations encountered in practice.

Numbers lend themselves just as readily to the purposes of cryptography as do letters, and many persons prefer numbers over letters, especially in transmitting messages to and from small points in the minor foreign countries where the Roman letters are not so easily read or written by native telegraph operators as are Arabic numerals, which are practically in use everywhere.

Number ciphers, just as letter ciphers, may be mono- or polyalphabetical in nature. A frequent type of number cipher is that

involving the use of the series of numbers from 00 to 99, making four complete alphabets of 25 letters each, thus allowing four values which may be chosen at random for each letter. Moreover, a key word can be chosen to designate the letter which begins each alphabet. Thus, the key word BLUE would indicate that the series 00 to 25 begins with 00 = B_p; 25 to 50 with 25 = L_p, etc., Figure 30.

00-B	25-L	50-U	75-E
01-C	26-M	51-V	76-F
02-D	27-N	52-W	77-G
03-E	28-O	53-X	78-H
04-F	29-P	54-Y	79-I-J
05-G	30-Q	55-Z	80-K
06-H	31-R	56-A	81-L
07-I-J	32-S	57-B	82-M
08-K	33-T	58-C	83-N
09-L	34-U	59-D	84-O
10-M	35-V	60-E	85-P
11-N	36-W	61-F	86-Q
12-O	37-X	62-G	87-R
13-P	38-Y	63-H	88-S
14-Q	39-Z	64-I-J	89-T
15-R	40-A	65-K	90-U
16-S	41-B	66-L	91-V
17-T	42-C	67-M	92-W
18-U	43-D	68-N	93-X
19-V	44-E	69-O	94-Y
20-W	45-F	70-P	95-Z
21-X	46-G	71-Q	96-A
22-Y	47-H	72-R	97-B
23-Z	48-I-J	73-S	98-C
24-A	49-K	74-T	99-D

FIGURE 30.

The four values given for each letter may be used at random, though, of course, they may just as easily be used in a simple multiple alphabet manner in the order 1-2-3-4. The solution of a cipher of either of these two types should present no difficulty, especially if the sequence of letters is the normal throughout. It is obvious, of course, that the sequences may be mixed sequences instead of the normal.

Another way of preparing a set of numerical values is to employ a diagram of the type shown in the accompanying Figure 31.

Here each letter may be represented by any one of three values

7261853094
362 ABCDEFGHIJ
815 KLMNOPQRST
794 UVWXYZ

FIGURE 31.

selected at random. A_p may be represented by 27, or 37, or 67; B_p by 22, 32, or 62, and so on. By arranging the numbers

at the sides of the rectangle in various ways, new series of values result. It is to be noted that the several values for a given letter agree in the second digit, and this fact affords immediate clues to the arrangement of the numbers at the sides of the rectangle. Here, too, the sequence of letters within the rectangle may be a mixed arrangement instead of the normal.

A method employed in 1915 by the Hindus in an attempt to stir up a rebellion in India made use of the rectangle shown in the accompanying Figure 32. After simple substitution, there were added the values of the letters of a key word, thus resulting in a typical multiple alphabet system. One of the keywords used was LAMP, whose value is

1234567
1 ABCDEFG
2 HIJKLMN
3 OPQRSTU
4 VWXYZ

L A M P
25-11-26-32.

FIGURE 32.

Message—Y O U R L E T T E R etc.
 Simple substitution—44 31 37 34 25 15 36 36 15 34 etc.
 Key values—25 11 26 32 25 11 26 32 25 11 etc.
 Cipher—69 42 63 66 50 26 62 68 40 45 etc.

The message was sent in an unbroken sequence of figures: 694263665026 . . . The solution of such a message should be apparent to the student. Repetitions of digraphs, trigraphs, and polygraphs disclose the length of the key, the text can be regrouped into individual alphabets, and those alphabets solved in the usual manner.

58. Digraphic substitution.—In all of the ciphers we have heretofore considered, substitution was monographic in nature; that is, the letters of the plain text were treated singly or individually in encipherment, resulting in a letter-for-letter substitution. In certain methods, however, instead of treating the letters singly, they are treated in pairs, or digraphs, thus giving rise to the term *digraphic substitution*. While digraphs, of course, have characteristic frequencies, still these frequencies are not so pronounced as in the case of individual letters, and, moreover, instead of dealing with only 26 frequencies, as in the case of monographic substitution, the cryptanalyst must consider 26×26 , or 676, frequencies in the case of digraphic substitution. For this reason digraphic substitution methods are in some respects somewhat more difficult to analyze than monographic methods, but in general the more complex types of polyalphabet methods are solved with greater difficulty than the usual types of digraphic substitution methods.

One of such digraphic methods employs a four-section square of the type shown in Figure 33. The plain-text pairs are taken from rectangles 1 and 2, and their equivalents are taken from rectangles 3 and 4, in such a manner that the plain-text pair and the equivalent

	W	A	S	H	I	A	B	C	D	E	
	N	G	T	O	B	F	G	H	I	K	
1	C	D	E	F	K	L	M	N	O	P	3
	L	M	P	Q	R	Q	R	S	T	U	
	U	V	X	Y	Z	V	W	X	Y	Z	
	A	B	C	D	E	L	I	N	C	O	
4	F	G	H	I	K	A	B	D	E	F	2
	L	M	N	O	P	G	H	K	M	P	
	Q	R	S	T	U	Q	R	S	T	U	
	V	W	X	Y	Z	V	W	X	Y	Z	

FIGURE 33.

cipher pair form the corners of a rectangle. Thus, $AB_p = BG_o$; $OR_p = GT_o$, etc. The letter J is omitted, I serving for its use. Note that rectangles 1 and 2 are based upon key words, and the secrecy of the method lies in frequent change of keys. Solution of this type of square would present no great difficulties if a single message of fair length or a series of messages enciphered by the same square are at hand. A tentative decipherment of portions of the cipher text based upon the frequency of digraphs would go hand in hand with the reconstruction of the enciphering square.

cipher pair form the corners of a rectangle. Thus, $AB_p = BG_o$; $OR_p = GT_o$, etc. The letter J is omitted, I serving for its use. Note that rectangles 1 and 2 are based upon key words, and the secrecy of the method lies in frequent change of keys. Solution of this type of square would present no great difficulties if a single message of fair length or a series of messages enciphered by the same square are at hand. A tentative

58a. **Playfair cipher.**—A development of the foregoing square is the *Playfair cipher* used by the British Army for a number of years, and to a limited extent by the United States Army during the late war. This consists of but one square; in the original system a key-word alphabet was inscribed in the square, beginning at the top and writing the letters from left to right in the ordinary manner; in the modified system, the letters within the square are in a mixed order, disarranged by shifting the columns and rows according to a systematic manner or by writing the letters in a more or less irregular manner in the square. The letter J is omitted, I serving for its use. The method of encipherment may be reduced to three cases. Let the square be as follows:

W	A	S	H	I
N	G	T	O	B
C	D	E	F	K
L	M	P	Q	R
U	V	X	Y	Z

(a) The pair of plain-text letters occupy diagonally opposite corners of a rectangle. Their equivalents are the letters at the other corners of the same rectangle, taken in the same order. Examples:

$$AR_p = IM_c; IM_p = AR_c; RA_p = MI_c; MI_p = RA_c.$$

(b) The pair of plain-text letters are situated in the same column. Their equivalents are the letters immediately below each of them in the same column and the bottom letter is represented by the top letter in that column. Examples:

$$AG_p = GD_c; DA_p = MG_c; VA_p = AG_c; AV_p = GA_c; VD_p = AM_c.$$

(c) The pair of plain-text letters are situated in the same horizontal row. Their equivalents are the letters immediately to the right of each of them in the same row and the letter on the extreme right is represented by the one at the extreme left in the same row. Examples:

$$CE_p = DF_c; CK_p = DC_c; FE_p = KF_c; EF_p = FK_c; KE_p = CF_c.$$

In encipherment, the plain text is first divided into pairs. If by chance a pair consists of a repeated letter, this pair is split by inserting a null or nonsignificant letter such as X, Y, or Z. Thus,

the word SUPPOSED, which would normally divide up into the pairs SU PP OS ED, is instead divided up into the pairs SU PX PO SE D-. The cipher text is, of course, transmitted in groups of five letters as usual.

The solution of such a cipher is not difficult, and, in cases where plenty of text is available, quite easy. Repetitions are bound to occur, and a frequency table of digraphs will permit of assumptions for such frequently occurring digraphs as ER, TH, IT, TO, etc. A tentative decipherment of a few such repetitions, accompanied by a reconstruction of the square will result in solution.¹ On account of the advances made in the field of cryptanalysis in recent years it is not thought that the Playfair cipher will be used in the future by the armies of the larger governments.

59. Trigraphic and polygraphic substitution.—It is of course possible to substitute trigraphically by means of tables, or even by means of sliding components, and such methods yield cryptograms highly difficult of analysis, even when an abundance of text is available. However, such methods are not suitable for field use on account of their impracticability. They were often used in diplomatic correspondence in the eighteenth century.

60. Machine ciphers.—It is a curious twist in human psychology which leads persons who have only the most rudimentary knowledge of cryptography and cryptanalysis, but who are otherwise apparently normal, to spend much money and labor in inventing and building "indecipherable" cipher machines. It is self-evident that the prerequisite to the invention of an "indecipherable" machine or system is a thorough knowledge of the science of cryptanalytics, and yet this obvious truth is neglected every day by would-be inventors of cipher machines. The cryptograms produced by many cipher machines can be analyzed very readily, because they can be reduced to terms of simple monoalphabet or polyalphabet ciphers, which are solvable by the ordinary processes outlined in the preceding pages. The reason for this is that machines must after all be mechanical in the cryptographic operations they perform, and this produces cyclic phenomena in the cryptograms.

There are only a few machine ciphers which may be considered to be reliable and small enough for military use. None of those which have been examined by the writer are "absolutely indecipherable." The cryptograms produced by certain large machines of an electrical nature are extremely difficult to solve, but these machines are entirely unsuited to field operations because of their bulk and their complexity,

¹ For detailed explanation of solution see Hitt's Manual and Mauborgne, Maj. J. O., *An Advanced Problem in Cryptography and its Solution*. Leavenworth Press, Fort Leavenworth, Kansas, 1914. Also see Langlois, André, *Cryptography*, E. P. Dutton and Company, 1922, pp. 166-188, for a solution by Comdr. W. W. Smith, U. S. N.

which necessitates special personnel. Furthermore, it must always be remembered that any machines used in the field are subject to capture, and therefore the enemy soon comes into possession of one. This enables him to analyze the machine, and thus gain an exact knowledge of the cryptographic operations executed by it.

As a demonstration of the fact that the complexity of a cryptographic machine, or the fact that it affords "millions of combinations" (to use a phrase so often employed by inventors of cipher machines), may have no direct bearing upon the complexity of the cryptograms produced by the machine, the following case is presented. Five short messages produced by a cryptographic machine for which emphatic claims of indecipherability were made were submitted for analysis. Here is the longest of the messages:

IUYHO	EWOQL	ZUUZQ	EKPOD	UAODX	RHOQF	PBMGG
SOJXV	X-TNW	-UYIT	BMCTC	ZLGTG	Y-PDS	QLZCN
AC-KE	TRODH	FCJAZ	HGTDV	MCGVH	X-DYR	LXYKQ
-VPTL	NIAHK	ULWGL	-RQLS	CNACY	PQMLX	IRYTU
QRDQF	LBXEU	EMXNG	-GOYA	FPGRU	OXIYS	PUBQJ
SEAPO	IZZFP	WVDAW	HDOEJ	QSYPQ	FGPVU	EVGIN
R-XMR	FDEDH	ERUTW	ILXMK	KBDX-	MYPOW	-PWVQ
SFJX-	BMIBD	UQFOD	VJZLQ	RTAX-	XNPIQ	FHAPP
FUHWI	TB-LS	IORJD	IHWAQ	TWABS	BAXQK	H-AXR
KDVLB	GVMKS	GFMMD				

Knowing that the machine was one of the typewriter style, and therefore in all probability employed the standard keyboard, a set of strips each bearing the sequence of the type bars of an ordinary three bank keyboard was made. The letter keys on the standard keyboard typewriter are arranged as follows:

Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M

The type bars (of letters) are arranged in the sequence QAZWSXEDC RFVGTGBYHNUJMIKOLP. The cipher, however, also showed an additional character "-", which meant that, in all probability, words were separated by this sign. It was necessary therefore to include this sign in the sequence of letters on the strips, and it was placed between B and Y, since these two letters are exactly at the midpoint of the upper and lower tiers of keys (including punctuation keys). The sliding strips were then arranged in the usual manner for a completion sequence solution and the various generatrices

studied for plain text. Note the diagram below, in which the plain-text letters have been underlined:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	5	6	7	8	9
<u>I</u>	<u>U</u>	<u>Y</u>	<u>H</u>	<u>O</u>	<u>E</u>	<u>W</u>	<u>O</u>	<u>Q</u>	<u>L</u>	<u>Z</u>	<u>U</u>	<u>Z</u>	<u>Q</u>	<u>E</u>	<u>K</u>	<u>P</u>	<u>O</u>	<u>D</u>	<u>U</u>	<u>A</u>	<u>O</u>	<u>D</u>	
K	J	H	N	L	D	S	L	A	P	W	J	J	W	A	D	O	Q	L	C	J	Z	L	C
O	M	N	U	P	C	X	P	Z	Q	S	M	M	S	Z	C	L	A	P	R	M	W	P	R
L	I	U	J	Q	R	E	Q	W	A	X	I	I	X	W	R	P	Z	Q	F	I	S	Q	F
P	K	J	M	A	F	D	A	S	Z	E	K	K	E	S	F	Q	W	A	V	K	X	A	V
Q	O	M	I	Z	V	C	Z	X	W	D	O	O	D	X	V	A	S	Z	T	O	E	Z	T
A	L	I	K	W	T	R	W	E	S	C	L	L	C	E	T	Z	X	W	G	L	D	W	G
Z	P	K	O	S	G	F	S	D	X	R	P	P	R	D	G	W	E	S	B	P	C	S	B
W	Q	O	L	X	B	V	X	C	E	F	Q	Q	F	C	B	S	D	X	-	Q	R	X	-
S	A	L	P	E	-	T	E	R	D	V	A	A	V	R	-	X	C	E	<u>Y</u>	<u>A</u>	<u>F</u>	<u>E</u>	<u>Y</u>
X	Z	P	Q	D	Y	G	D	F	C	T	Z	Z	T	F	Y	<u>E</u>	<u>R</u>	<u>D</u>	<u>H</u>	<u>Z</u>	<u>V</u>	<u>D</u>	<u>H</u>
E	W	Q	A	C	H	B	C	V	R	G	W	W	G	V	<u>H</u>	<u>D</u>	<u>F</u>	<u>C</u>	<u>N</u>	<u>W</u>	<u>T</u>	<u>C</u>	<u>N</u>
D	S	A	Z	R	N	-	R	T	F	B	S	S	B	<u>T</u>	<u>N</u>	<u>C</u>	<u>V</u>	<u>R</u>	<u>U</u>	<u>S</u>	<u>G</u>	<u>R</u>	<u>U</u>
C	X	Z	W	F	U	Y	F	G	V	-	X	X	-	<u>G</u>	<u>U</u>	<u>R</u>	<u>T</u>	<u>F</u>	<u>J</u>	<u>X</u>	<u>B</u>	<u>F</u>	<u>J</u>
R	E	W	S	V	J	H	V	B	T	<u>E</u>	<u>E</u>	<u>Y</u>	<u>B</u>	<u>J</u>	<u>F</u>	<u>G</u>	<u>V</u>	<u>M</u>	<u>E</u>	-	<u>V</u>	<u>M</u>	<u>M</u>
F	D	S	X	T	M	N	T	-	<u>G</u>	<u>H</u>	<u>D</u>	<u>D</u>	<u>H</u>	-	<u>M</u>	<u>V</u>	<u>B</u>	<u>T</u>	<u>I</u>	<u>D</u>	<u>Y</u>	<u>T</u>	<u>I</u>
V	C	X	<u>E</u>	<u>G</u>	<u>I</u>	<u>U</u>	<u>G</u>	<u>Y</u>	<u>B</u>	<u>N</u>	<u>C</u>	<u>C</u>	<u>N</u>	<u>Y</u>	<u>I</u>	<u>T</u>	-	<u>G</u>	<u>K</u>	<u>C</u>	<u>H</u>	<u>G</u>	<u>K</u>
T	<u>R</u>	<u>E</u>	<u>D</u>	<u>B</u>	<u>K</u>	<u>J</u>	<u>B</u>	<u>H</u>	-	<u>U</u>	<u>R</u>	<u>R</u>	<u>U</u>	<u>H</u>	<u>K</u>	<u>G</u>	<u>Y</u>	<u>B</u>	<u>O</u>	<u>R</u>	<u>N</u>	<u>B</u>	<u>O</u>
<u>G</u>	<u>F</u>	<u>D</u>	<u>C</u>	-	<u>O</u>	<u>M</u>	-	<u>N</u>	<u>Y</u>	<u>J</u>	<u>F</u>	<u>F</u>	<u>J</u>	<u>N</u>	<u>O</u>	<u>B</u>	<u>H</u>	-	<u>L</u>	<u>F</u>	<u>U</u>	-	<u>L</u>
B	V	C	R	Y	L	I	Y	U	H	M	V	V	M	U	L	-	<u>N</u>	<u>Y</u>	<u>P</u>	<u>V</u>	<u>J</u>	<u>Y</u>	<u>P</u>
-	T	R	F	H	P	K	H	J	N	I	T	T	I	J	P	Y	U	H	Q	T	M	H	Q
Y	G	F	V	N	Q	O	N	M	U	K	G	G	K	M	Q	H	J	N	A	G	I	N	A
H	B	V	T	U	A	L	U	I	J	O	B	B	O	I	A	N	M	U	Z	B	K	U	Z
N	-	T	G	J	Z	P	J	K	M	L	-	-	L	K	Z	U	I	J	W	-	O	J	W
U	Y	G	B	M	W	Q	M	O	I	P	Y	Y	P	O	W	J	K	M	S	Y	L	M	S
J	H	B	-	I	S	A	I	L	K	Q	H	H	Q	L	S	M	O	I	X	H	P	I	X
M	N	-	Y	K	X	Z	K	P	O	A	N	N	A	P	X	I	L	K	E	N	Q	K	E

Note also that the "route" formed by underlining successive plain-text letters constitutes a cycle of 15 units, the cycle of "steps" being repeated over and over again throughout the message. The elements of the cycle consist merely in whether the second letter of a pair is (1) on the same generatrix with the first letter of that pair, (2) on the generatrix above it, or (3) on the generatrix below it. The "millions of keys" of the system merely determine (1) the starting point and (2) the "steps" in the cycle. We see here that the starting point is of absolutely no concern to the cryptanalyst, for he will pick up his first pair of letters on some pair of generatrices of the completion sequence diagram, and that the steps of the cycle constitute no stumbling block for him because he can pick up the intelligible text very quickly after a start is made.

The completely deciphered message is as follows:

Greeting thee. There are those who believe that ther (e are) weaknesses in this machine on account of the cycle movement. If this is so this matter should be sufficient to prove no contention. Try it out and advise us of your success. We do not believe that it is possible. At least we should like to see it proven. Go to it. Report at any old time.

It happened in this case that the machine employed a cylindrical type wheel instead of type bars (Blickensderfer type of machine) but the sequence of letters on the type wheel coincided with the sequence of type bars on the standard keyboard. Even if a type wheel with an unknown sequence of letters were substituted, a series of messages could be solved by superimposition of cycles. Note that sequent cycles can not be superimposed, for these will not produce columns in which the letters belong to the same single mixed alphabet. For example, the fourth letter of the first cycle of this message H_0 equals E_p , and the fourth letter of the second cycle O_0 also equals E_p . But the 1st cycle can be superimposed upon the 28th, 55th, 82d, and so on; the 2d cycle can be superimposed upon the 29th, 56th, 83d, and so on. The alphabet here consists of 27 characters, which accounts for the series 1, 28, 55 . . . , and 2, 29, 56 The student is urged to study the mechanics of this solution for it will give him a good insight into certain types of cyclic phenomena.

61. Concluding remarks on polyalphabet substitution ciphers.—As has already been mentioned, the various types of polyalphabet substitution ciphers illustrated in the foregoing pages by no means exhaust the category of such methods, but no comprehensive treatment of the more unusual systems can here be given. In general it may be said that whenever the text of such cryptograms, or a series of them, can be reduced to the simple elements of a single alphabet frequency distribution, or a series of single alphabet frequency distributions, the cryptograms may nearly always be solved. The most difficult part of the analysis usually involves this breaking up of the heterogeneous or mixed text of a plurality of different cipher alphabets into the homogeneous or pure elements of single cipher alphabet distributions. Where cyclic or periodic phenomena are manifested in the cryptograms this reduction into simple elements is very much easier than where no such phenomena can be found. Once the individual alphabets have been isolated, they can be analyzed on the basis of frequency. Where these alphabets are interrelated secondary alphabets produced by sliding components, solution is rendered somewhat easier by taking advantage of the principles of direct or indirect symmetry of position. Where the alphabets are independent or not related in any way, as in random mixed alphabets, the analysis is somewhat slower and more difficult because no aids can be brought to bear upon a solution by the simple principles of frequency.

SECTION XV.

TRANSPOSITION CIPHERS—STANDARD METHODS OF SOLUTION.

	Paragraph
General remarks	62
Completely filled rectangles—route transposition	63
Completely filled rectangles—columnar transposition	64
Solution of a typical case	64a
The Federal Army cipher	64b
Incompletely filled rectangles	65
Solution of a typical example	65a
Double transposition	66
Grilles and other devices	67

62. General remarks.—Transposition ciphers, as stated before, involve no change in the identity of the letters of the plain text; only their original order or arrangement is changed. Sometimes only the order of the words is changed, as in the so-called "Route Cipher" employed by the Federal Army. We may call the latter type of transposition *word transposition*, using the term *monoliteral transposition* to apply to the former type, in which the individual letters are rearranged. The student who has carefully followed the principles of solving the common types of substitution ciphers and has now become practiced in the handling of cryptograms so that he no longer experiences the perplexity and bewilderment that first seized him when he began his study of cryptography, will have no difficulty with the ordinary types of transposition ciphers. The most commonly encountered types are for the most part simpler to analyze than the more difficult types of substitution ciphers.

Practically all monoliteral transposition ciphers involve the use of a design or a geometrical figure, usually a rectangle, in which the plain-text letters are written in one way and taken out to form the letters of the cipher text in some other way. In nearly all cases the secrecy of the method consists in employing rectangles of varying dimensions and varying the order in which the letters are put in and taken out, every operation usually being governed by a key.

For our purposes, we may arbitrarily classify the usual designs or rectangles into two types: (1) Those in which the rectangle is symmetrical, or completely "filled," i. e., in which there are no vacant spaces that might be occupied by letters but are not so occupied; and (2) those in which the rectangle is not symmetrical, or completely "filled." The reason for this classification will become apparent in the subsequent discussion.

63. Completely filled rectangles—route transposition.—Suppose the correspondents agree to use a rectangle of eight columns. The message is inscribed within the rectangle, and if a few vacant spaces are left at the end, nulls or "dummy" letters are inserted so as to complete the rectangle. Then one of many different paths or *routes* are followed in rewriting the letters to produce the transposed text, and it is possible for each route to have a different starting point, this also being agreed upon by the correspondents. Normally the starting point is one of the four outside corners of the rectangle. In his *Manual for the Solution of Military Ciphers*, Colonel Hitt gives the routes shown in Figure 35. According to route *h*, starting at the upper left hand corner, the message shown above would become AAODO TWOAM YZWTE PHKCA TTSSU MORRO etc.

A	T	T	A	C	K	H
A	S	B	E	E	N	P
O	S	T	P	O	N	E
D	U	N	T	I	L	T
O	M	O	R	R	O	W
T	W	O	A	M	Y	Z

FIGURE 34.

(a) Simple horizontal:

ABCDEF FEDCBA STUVWX XWVUTS
GHIJKL LKJIHG MNOPQR RQPONM
MNOPQR RQPONM GHIJKL LKJIHG
STUVWX XWVUTS ABCDEF FEDCBA

(b) Simple vertical:

AEIMQU DHLPTX UQMIEA XTPLHD
BFJNRV CGKOSW VRNJFB WSOKGC
CGKOSW BFJNRV WSOKGC VRNJFB
DHLPTX AEIMQU XTPLHD UQMIEA

(c) Alternate horizontal:

ABCDEF FEDCBA XWVUTS STUVWX
LKJIHG GHIJKL MNOPQR RQPONM
MNOPQR RQPONM LKJIHG GHIJKL
XWVUTS STUVWX ABCDEF FEDCBA

(d) Alternate vertical:

AHIPQX DELMTU XQPIHA UTMLED
BGJORW CFKNSV WROJGB VSNKFC
CFKNSV BGJORW VSNKFC WROJGB
DELMTU AHIPQX UTMLED XQPIHA

(e) Simple diagonal:

ABDGKO GKOSVX OKGDBA XVSOKG
CEHLPS DHLPTW SPLHEC WTPLHD
FIMQTV BEIMQU VTQMIF UQMIEB
JNRUWX ACFJNR XWURNJ RNJFCA

(f) Alternate diagonal:

ABFGNO GNOUVX ONGFBA XVUONG
CEHMPU FHMPW UPMHEC WTPMHF
DILQTV BEILQS VTQLID SQLIEB
JKRSWX ACDJKR XWSRKJ RKJDCA

ACFJNR JNRUWX RNJFCA XWURNJ
BEIMQU FIMQTV UQMIEB VTQMIF
DHLPTW CEHLPS WTPLHD SPLHEC
GKOSVX ABDGKO XVSOKG OKGDBA

ACDJKR JKRSWX RKJDCA XWSRKJ
BEILQS DILQTV SQLIEB VTQLID
FHMPW CEHMPU WTPMHF UPMHEC
GNOUVX ABFGNO XVUONG ONGFBA

(g) Spiral, clockwise:

ABCDEF LMNOPA IJKLMNOP DEFCHI
PQRSTG KWXQB HUVWYO CRSTUJ
OXWVUH JUTSRC GTSRQP BQXWVK
NMLKJI IHGFED FEDCBA APONML

(h) Spiral, counterclockwise:

APONML NMLKJI IHGFED FEDCBA
BQXWVK OXWVUH JUTSRC GTSRQP
CRSTUJ PQRSTG KWXQB HUVWYO
DEFCHI ABCDEF LMNOPA IJKLMNOP

FIGURE 35.

It is apparent that instead of writing the plain-text letters within the rectangle in the normal manner, i. e., from left to right and from the top downwards, the letters may be inscribed according to any one of the routes agreed upon and then the cipher text written out in the usual five letter groups by taking the letters from the rectangle in the normal manner, i. e., in this case from left to right, and from the top downwards, or by following another route of transposition.

The solution of such cryptograms¹ is merely a matter of experimenting with rectangles of various dimensions suggested by the total number of letters in the message, and then inspecting these rectangles, searching for portions of words by reading horizontally, diagonally, vertically, spirally, and so on. In many of the routes no experiment is even necessary to determine the dimensions of the rectangle in order to get a starting point. Consider the following example:

MESSAGE.

LEOCO ADPIG HHNSI NAGWI IUNLO TDGTV HOLES
MYHSF DRARN OSIEI EANVV OUNEE GSTEN NPNAB EC

Suppose we start off by assuming an alternate diagonal route. The first three groups yield text at once, the word LEAD or LEADING becoming apparent in the first line, and TROOPS in the second line. (Fig. 36.) We continue the method until the sense breaks, as shown in Figure 37, when we have reached the group MYHSF, with columns of nine letters assumed. We therefore assume columns of eight letters and go back to the point where the sense broke. It is seen that the alternate diagonal route was the one followed in encipherment, and that a rectangle 9×8 was used. (Fig. 38.) The three letters A B C are nulls inserted to complete the rectangle. Of course the cryptanalyst might not hit upon the method until after he had tried several other routes, but such trials take only a very few minutes. It is to be noted that large rectangles of this type are not often encountered because in many of the routes whole plain-text words show up in the cipher text, and also because it would often be necessary to insert many nulls to fill up the last line of the rectangle if the message is but a few letters longer than an exact multiple of the number of columns.

LEADI
OOPS
CIN
GH
H

FIGURE 36.

LEADINGTN	LEADINGTR
OOPSADVR	OOPSADVAN
CINGTHA	CINGTHROU
GHWOOR	GHWOODSON
HILLD	HILLFIVEN
INEF	INESEVENP
USS	USHINGENE
MH	MYEASTABC
Y	

FIGURE 37.

FIGURE 38.

¹ For a detailed treatment of ciphers of this type see under bibliography, *Riverbank Publication No. 19*, by Capt. Lenox R. Lohr.

64. **Completely filled rectangles—columnar transposition.**—One of the very common types of transposition is that in which the vertical columns, or the horizontal rows, or both, are rearranged according to a key number. Such a key number, or *numerical key*, as it is termed, is most often derived from a word or a phrase, by assigning the numbers 1, 2, 3, . . . to the letters according to their relative positions in the normal alphabet. For example, given the key word WILMINGTON the numerical key is derived as follows:

W - I - L - M - I - N - G - T - O - N
10 - 2 - 4 - 5 - 3 - 6 - 1 - 9 - 8 - 7

It is to be noted that similar letters in the key word are assigned numbers in sequence, from left to right, as they appear in the word.

The text of the message is written in regular lines beneath the numerical key, thus forming a rectangle, which, if necessary, is completely filled by the addition of dummy letters. In encipherment the letters are taken from the columns as read in numerical order and sent in groups of five letters. Example:

Key word—	L I B E R T Y	
Numerical key—	4 3 1 2 5 6 7	
	A T T A C K P	
	O S T P O N E	TTNAA PTMTS UOAO
	D U N T I L T	WCOIX KNLYP ETZ
	W O A M X Y Z	

FIGURE 39.

64a. **Solution of a typical case.**—In solving a cryptogram of this type the cryptanalyst finds clues to the dimensions of the rectangle from the factors of the total number of letters. He then writes the message in a rectangle of the most probable dimensions in columnar fashion, cuts the columns apart, and tries to rearrange them in proper order to give intelligible text. This process of matching columns in order to build up words is termed ANAGRAMMING. The method is to select a column which has a good assortment of high-frequency letters and try to find columns which can be added before and after the selected column to build up high-frequency digraphs and trigraphs. Once a set of three or four columns has been correctly assembled, it is very easy to complete the process. An example follows.

MESSAGE.

VEDEJ OBEEN MTOVE LOHXD WRUOE THBSS WTEDS IVTLE
 NRSGR OIORU ANEBR RMEYO LTWTT DIEOK HETGI AAROR
 UFUCD MORRA MLAEP RRPBI TEGNV VRCTC NEREA RETYT
 SSUOO YZDLI TTICE TERSY ELITB OAEWN ES

This message contains 152 letters, suggesting a rectangle 19 by 8. It may be 19 columns of 8 letters each, or 8 columns of 19 letters each. We write the message out according to both assumptions, as in Figure 40, and count the number of vowels in each horizontal line in order to see which is the more probable arrangement. Since normal English military text consists of approximately 40 per cent vowels, a line of 19 letters should contain 7.6 vowels and a line of 8 letters, 3.2 vowels.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	Vow- e.s.	1 2 3 4 5 6 7 8	Vow- els.
VEOENRMTTURRVSLEB	6	VDLEARAC	3
ENHTDRUEDGFAPVESIRO	7	EWYRRRE	4
DMXHSSAYIIUMBRAUTSA	8	DRNOOPET	3
ETDBIGNOEACLICROTYE	9	EURLRBTE	3
JOWSVRELOADATTEOIEW	9	JOSTUIYR	4
OVRSTOBTKRMEECTYCLN	5	OEGWFTTS	2
BEUWLIRWHOOPGNYZEIE	9	BTRTUESY	3
ELOTEORTERRRNETDTTS	6	EHOTCGSE	3
		EBIDDNUL	3
		NSOIMVOI	4
		MSREOVOT	3
		TWUORRYB	3
		OTAKRCZO	3
		VENHATDA	3
		EDEEMCLE	3
		LSBTLNIW	1
		OIRGAETN	4
		HVRIERTE	3
		XTMAPEIS	3

A

B

Average deviation from theoretical
vowel content.

Arrangement A:

Total deviation -11.0
 No. of lines - 8.
 Average deviation= 1.38

Arrangement B:

Total deviation - 9.2
 No. of lines -19.
 Average deviation= .48

FIGURE 40.

The arrangement of 8 columns gives on the whole a smaller deviation in vowel content from the theoretically expected content in the various lines than does the arrangement of 19 columns, and we select the former arrangement as the more probable. The columns of this arrangement are now cut apart and the anagramming process is applied to them.

In some languages there are two letters, each of low individual frequency, which when present in a piece of text are always combined as a pair, thus forming a digraph of medium frequency. The letters C and H in German, for example, behave in this manner, as do the letters Q and U in French, Spanish, and Italian.

Even a short message in German will contain one or two C's and H's, which the cryptanalyst knows must form the invariable combination CH. In English, also, the digraph QU is an invariable combination, but it is of very low frequency, and is therefore not often encountered in short messages. Furthermore, in military messages even if a Q does occur, it does not follow absolutely that it is a part of the digraph QU, for abbreviations such as SQ for "squadron," REQ for "request," etc., are common, and Q is also often used as a null. There are no other invariable digraphs in this language, so that the anagramming process can not make great use of such invariable combinations in order to get a start. The nearest approach to the condition of an invariable combination in English is TH, and we have seen that this digraph, which is of greatest frequency in ordinary literary text, is of low frequency in telegraphic text. Moreover, T is a letter of high frequency and H a letter of medium frequency, so that in a single message there may be many T's and a few H's, so that the exact letters which go together to form the one or two TH's in a message are not so easily determined.

However, we do not really need to have an invariable combination with which to obtain a start in anagramming. We may use the ordinary high-frequency digraphs ER, ON, IN, ES, EN, AT, AN, etc., as a basis.

Returning now to our problem, we examine the columns of Figure 40 B, in order to find one which has the most high-frequency letters, and then search for columns which can be added to it to yield

7-1-5

A V A

R E R

E D O

T E R

Y J U

T O F

S B U

S E C

U E D

O N M

O M O

Y T R

Z O R

D V A

L E M

I L L

T O A

T H E

I X P

FIGURE 41.

FIGURE 42.

good digraphs and trigraphs in the lines thus formed. Column 1 is as good as any of the others for a start.

We experiment with all the other columns in 6-4-2-8-7-1-5-3 juxtaposition with column 1, and finally obtain R E D C A V A L the combination 7-1-5 (fig. 41), which seems R Y W E R E R E to give very good trigraphs. In fact, with a P O R T E D O N little imagination words immediately suggest B L U E T E R R themselves: AVA suggests CAVALRY; SEC suggests I T O R Y J U S SECOND; ILL suggests WILL; etc. If T W E S T O F G we add to these three columns such columns E T T Y S B U R as will produce the word CAVALRY, suggested G T H E S E C O by AVA, we obtain the arrangement shown in N D B L U E D I Figure 42. There is no doubt about the correct- V I S I O N M O ness of the arrangement shown in this figure, for V E S T O M O R whole words now stand out prominently. It is R O W B Y T R U easy to complete the rectangle, which is shown C K T O Z O R A in Figure 43. Only a few minutes are necessary T H E A D V A N to solve a cryptogram of this type, and it is C E D E L E M E obvious that such a method of encipherment N T S W I L L B affords no security whatsoever against rapid solu- E G I N T O A R tion by the enemy. R I V E T H E R

Often it is found that after the columns have E A T S I X P M apparently been properly assembled so as to yield FIGURE 43. intelligible words on individual lines the message still does not read consecutively from top to bottom. Here the rows have been rearranged as well as the columns, a system which is said to have been used by the Russian Nihilists. Such a twofold transposition does not greatly complicate solution.

64b. The Federal Army cipher.—Instead of transposing individual letters, it is of course possible, as has already been stated, to transpose entire words and employ any of the routes indicated on page 94, or the columnar transposition method. In fact, the so-called Federal Army cipher, used extensively by the Union Army during the Civil War, was this type of cipher, and if one can place much credence in the meager reports available, the Confederates were unable to solve the messages, even going to the lengths of advertising in the newspapers for cipher experts skilled in "deciphering without the key." No cipher of this type would afford any security to-day. In its original form only one route was followed, going down certain

columns and up other columns, the columns being taken in a mixed order. Also, nulls or blind words having no significance in the message were inserted; often words were intentionally misspelled, such as "meat" for "meet," "wood" for "would," etc., and also conventional or code names were used for the names of persons and places. In this connection it is historically interesting to give the following message, supposedly sent by President Lincoln to Major General Burnside:

WASHINGTON, November 25, 1862.

BURNSIDE, *Falmouth, Virginia.*

Can Inn Ale me with 2 oar our Annpas Ann me flesh ends N. V. Corn Inn out with U and Inn Heaven day nest Wed roe Moore Tom darkey hat greek a Why Hawk of Abbott Inn B chewed I if.

BATES.

Reading the message backwards and correcting spelling the "cryptogram" reads:

WASHINGTON, November 25, 1862.

BURNSIDE, *Falmouth, Virginia.*

If I should be in a boat off Aguia Creek at dark tomorrow, Wednesday evening, could you, without inconvenience, meet me and pass an hour or two with me? A. Lincoln.

BATES.

65. Incompletely filled rectangles.—One of the reasons why the solution of transposition ciphers based upon completely filled rectangles is so easy is that the columns are all of the same length, so that the cryptanalyst has no difficulty in anagramming. This is so for the reason that the anagramming process is applied to letters which really come from the same line of the enciphering diagram. If the columns were not all of the same length, then the cryptanalyst would not be sure that the letters which he is attempting to anagram really do come from the same line. By making sure that the last line in the rectangle is never completely filled, and applying a columnar transposition as usual, the encipherer thus prevents the cryptanalyst from being able to divide up the cipher text into columns of equal length. Here is a short example of encipherment.

Key word—S C I E N C E	
Numerical key—7-1-5-3-6-2-4	
A T T A C K P	MESSAGE.
O S T P O N E	TSURC KNLCA PTEPE
D U N T I L T	TLTTN EKCOI OAODH O
H R E E O C L	
O C K	

FIGURE 44.

65a. Solution of a typical example.—The solution of such a cryptogram is, however, only a little more difficult than that in which the columns are all of equal length, produced by completely filling the rectangle.

MESSAGE.

EPAHR ELNFO CPOIO ASROS MWIWA SOTAV TIOES
 INDED DXSTAT YNTU OMEPC ASFCR LTORN MOGRA
 SAORO SLLAH AMGAB OGTEW DOSME ILOOO DRFTA
 TVABT RHDER DEXPH NOYIP WIITO CEDON WIRTU
 LMRNX LWAIN OVTUH LIONN XARLU SOTSS RNNMI
 EGERH AATCE FRMOS NS

This message contains 192 letters. Since we are now dealing with incomplete rectangles it will be impossible to assume rectangles whose dimensions are suggested by the factors of 192 and divide the text up into columns of equal length. Considerable experiment is necessary to determine the exact size of the rectangle, and the various assumptions for size would have to be tested out in the same manner as the test now to be described, based upon an assumption of a rectangle of 18 columns.

If the "set up" contains 18 columns, then there must be 12 columns of 11 letters and 6 columns of 10 letters. Of course, we do not know which columns are the ones containing the extra letter at the bottom, but we can write the message in columns of 11 letters each, as far as possible, using cross section paper and leaving room at the top and bottom of the columns for the insertion of such letters as may be later necessary. (Fig. 45.)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
E	P	I	E	T	A	O	L	E	O	R	O	D	N	H	S	G	M
P	O	W	S	Y	S	G	A	W	D	H	Y	O	X	L	O	E	O
A	I	A	I	T	F	R	H	D	R	D	I	N	L	I	T	R	S
H	O	S	N	N	C	A	A	O	F	E	P	W	W	O	S	H	N
R	A	O	D	T	R	S	M	S	T	R	W	I	A	N	S	A	S
E	S	T	E	U	L	A	G	M	A	D	I	R	I	N	R	A	
L	R	A	D	O	T	O	A	E	T	E	I	T	N	X	N	T	
N	O	V	D	M	O	R	B	I	V	X	T	U	O	A	N	C	
F	S	T	X	E	R	O	O	L	A	P	O	L	V	R	M	E	
O	M	I	T	P	N	S	G	O	B	H	C	M	T	L	I	F	
C	W	O	A	C	M	L	T	O	T	N	E	R	U	U	E	R	

FIGURE 45.

We first look to see if by chance there is a Q in the message, which might be part of the digraph QU, but are disappointed. The digraph TH can not be relied upon, for this is a military message and THE is probably omitted. We do, however, note the letter V, and this, together with other circumstances, suggests that the word CAVALRY might be present. Let us anagram for this word. Columns 3, 10, and 14 each show a letter V, and in approximately the same position

in the columns. We now combine each of these columns containing a V with every column containing an A in order to form the trigraph

<p><u>14</u> N X L W <u>3</u> A <u>1</u> I I N E W O P A V A S T H O U R T H E A L L V I N T O F I N O O N C</p>	<p>AVA, and of course we must not restrict ourselves to those columns only which show the letter A on the same line with V, for we must always remember that our columns are not all the same length in the enciphering rectangle, and therefore it will probably be necessary to slide the columns several letters up or down in testing out various assumptions. After more or less experiment we have assembled the columns shown in Figure 46, and it is to be noted that we have found it necessary to add to column 14 some of the letters which our arbitrary grouping of the cipher text placed in column 15. All the trigraphs given by this combination are excellent. We note also that E, the first letter of the message is part of the trigraph INE, as shown in Figure 46. This means that the cipher text should be cut so that column 3</p>	<p>17 14 <u>18-3-15-1-8</u> A I N E S T W O P L C A V A L E S T H A F O U R H R T H E A M A L L M O V I N G S T O F A N I N O B S O N C O</p>
--	--	---

FIGURE 46.

FIGURE 47.

begins with letter I, and column 14 or 15 with letter N. Since the letter I under discussion is the 23d letter of the message, it follows that columns 1 and 2 are both "long" columns, i. e., they have 11 letters each, on the basis of a diagram of 18 columns. We continue to anagram for our assumed word CAVALRY. (Note fig. 47.)

Other words now become prominent: TWO PL(ATOONS?), SMALL, MOVING, etc. Following up these suggestions the rectangle is speedily completed and we have:

6 11 2 14 9 18 3 15 1 8 16 5 13 10 17 4 12 7
 C A P T G A I N E S X A T O N E P M
 A B O U T T W O P L A T O O N S H O
 S T I L E C A V A L R Y C O M I N G
 F R O M W E S T H A L T E D I N O R
 C H A R D F O U R H U N D R E D Y A
 R D S N O R T H E A S T O F G E I S
 L E R X S M A L L M O U N T E D P A
 T R O L M O V I N G T O W A R D W O
 O D S W E S T O F A S M I T H X I R
 R E M A I N I N O B S E R V A T I O
 N X W I L S O N C O R P

FIGURE 48.

It seems hardly necessary to say that the solution is after all a matter of experiment, trial of many assumptions, rejection of those which lead to no results, and establishment of a few columns which seem to give the best results. Once a start is obtained, the rest follows easily and rapidly.

66. Double transposition.—Simple as the solution of the foregoing type of transposition appears, the application of one more step in encipherment results in a transposition cipher which under certain conditions will defy solution. If after the first rectangle has been inscribed with the plain text, the letters of the columns, taken in key number order, are inscribed *horizontally* in a rectangle based upon the same key, from left to right and from the top downwards, and then the cipher text is taken from the *columns* of the second rectangle in key number order, a thorough rearrangement of letters is effected. An example of encipherment follows:

	Plain-text rectangle.	First transposition.	Second transposition.
Key word—	B U R E A U		
Key numbers—	2-5-4-3-1-6	2-5-4-3-1-6	
	A T T A C K	C P N U A P	MESSAGE.
	P O S T P O	N I A A T U	ATFKC NOOTU ADMNA
	N E D U N T	O T S D F T	SLPIT ERPUT O
	I L F O U R	O E L M K O	
	A M	T R	

FIGURE 49.

Although, theoretically, such a double transposition cipher is extremely difficult to solve, nevertheless there are certain disadvantages to the system when used in military cryptography. The greatest danger is the failure on the part of careless clerks to execute both transpositions. The interception and solution of a single message which has undergone but one transposition will immediately provide the key for the solution of all the other messages even though they have been correctly enciphered. Again, the interception of two or more messages of exactly the same length will provide material for solution, as will be explained below. Again, the interception of a message which is based upon a perfect square, even though both transpositions have been effected properly, will enable solution; and the solution of completely filled rectangles though they may not be perfect squares, is possible. It is therefore difficult to regulate the use of this method so as to preclude the possibility of decipherment, but nevertheless, the method is admittedly a very excellent one on account of its ease, simplicity, rapidity, and comparative safety when infrequently used.

67. Grilles and other devices.—A type of transposition occasionally used by secret agents, spies, and the like is that involving

the use of a square sheet of paper or cardboard, called a *grille*, in which small square perforations have been cut in definite but irregular positions, and the letters of the plain text are inscribed on a sheet underneath the perforated design. Usually, the grille is revolved 90° in four successive operations so that the resulting square of letters inscribed beneath the grille is completely filled, and then the letters are taken in groups of five, reading horizontally, or otherwise according to agreement. The perforations in the grille, must, of course, be correctly disposed on the grille so as to produce the result that every space on the sheet over which it is placed in inscribing the letters shall be filled after the four turns of the grille have been completed. Such ciphers are not difficult to solve, and where several messages of similar length have been intercepted, the general solution given below may be applied. The grille may then be reconstructed by an analysis and comparison of the cipher and the corresponding plain text.

Irregular designs of various types may be used, but they necessitate, as does the grille, the carrying about of such designs and devices. They are therefore not suitable for military cryptography, but are occasionally used in lieu of better methods. Grilles were used by the Germans during the early part of the war, but were soon discarded in favor of more scientific methods.

SECTION XVI.

TRANSPOSITION CIPHERS—SPECIAL METHODS OF SOLUTION.

	Paragraph.
Messages of identical length.....	68
Messages with similar beginnings.....	69
Messages with similar endings.....	70
Concluding remarks on transposition ciphers.....	71

68. Messages of identical length.—The essential feature of transposition ciphers lies in the alterations in the positions of the letters composing the plain-text message, the changes being determined by a key. *It follows, therefore, that the letters of two or more messages of exactly the same length, when enciphered by the same key, will undergo exactly the same alterations in positions.* Given several such messages, solution may be attained by writing them under one another and anagramming the *columns* thus formed. An example will serve to make the method clear, using the extremely difficult double transposition cipher mentioned above.

MESSAGES¹.

1. RJETU EBDIT TLVTR IRYCN EORHN EATSI OOSTF
CTDDR VFEAY MSNLD EOXER EHPYR UIEIE PIEMN AO
2. NTEIT EROML CNANC INUCM LELAE NLASP OAMSE
LTEON RTVOL TRITI YAXWT IEHSN EEFPR SPFII IS
3. AQELE EYDS EAEAN CMNAU ONCND AICRG NRGNA
ERTEA BCIPH ERTNC ATTAC USALN DTUHE IRPWM VB

These three messages are of exactly the same length and are suspected of being in the same key. We write them out in super-imposed lines thus:

1	5	10	15	20	25	30	35	40	45	50
1.	RJETUEBDITTLVTR	I	RYCNEORHNEATS	I	OOSTFCTDDR	V	F	EAYMSNLD		
2.	NTEITEROMLCNANC	I	NUCMLLELAENLAS	P	OAMSELTEONRT	V	OLTRITI			
3.	AQELEEEYDSEAEAN	C	MNAUONCND	A	ICRGNRGNAERTE	A	B	CIPHERTNC		
		55	60	65	70					

FIGURE 50.

In Message 3 we find a Q in column 2. There are three columns which show the letter U in Message 3, viz, 20, 56, and 63. Let us place column 2 in position with each of them:

2-20	2-56	2-63
J N	J E	J E
T M	T I	T F
Q U	Q U	Q U

FIGURE 51.

We may discard the combination 2-20 at once. Combination 2-63 is possible, but does not look as good as 2-56, for TI is a very frequent digraph. Let us try to add to the 2-56 combination by assuming that the QU is part of the word REQUEST. Here are the combinations possible for QUE:

2-56-3	2-56-5	2-56-6	2-56-7	2-56-11
J E E	J E U	J E E	J E B	J E T
T I E	T I T	T I E	T I R	T I C
Q U E	Q U E	Q U E	Q U E	Q U E
2-56-13	2-56-36	2-56-39	2-56-46	2-56-65
J E V	J E C	J E D	J E M	J E E
T I A	T I L	T I O	T I T	T I R
Q U E	Q U E	Q U E	Q U E	Q U E

FIGURE 52.

¹ As an instance of a most remarkable coincidence, note the appearance of the word CIPHER in the cipher text of the third message. Theoretically, such an event will happen, as a result of chance, once in 26⁴ (-308,915,776) times. The word CIPHER does not appear in the plain-text message at all!

Of these we select combination 2-56-36 as the best, for JEC suggests the word OBJECT or OBJECTIVE. We next try to add to our 2-56-36 combination columns which will give OBJECT for Message 1 and REQUEST for Message 3. Note what we have from this combination:

<u>32-7-2-56-36-10</u>						
O	B	J	E	C	T	
A	R	T	I	L	L	
R	E	Q	U	E	S	

FIGURE 53.

The word ARTILLERY stands out. Let us add to the three words already appearing. We search for a column which has T in the third line, E in the second, and possibly I in the first. Column 62 fills these conditions:

<u>32-7-2-56-36-10-62</u>						
O	B	J	E	C	T	I
A	R	T	I	L	L	E
R	E	Q	U	E	S	T

FIGURE 54.

Next, a column showing V and R in lines 1 and 2, respectively, is sought. Column 41 fills the conditions:

<u>32-7-2-56-36-10-62-41</u>							
O	B	J	E	C	T	I	V
A	R	T	I	L	L	E	R
R	E	Q	U	E	S	T	B

FIGURE 55.

Further progress is made in similar manner until the messages are completely deciphered. The key for the transposition may now be reconstructed by an analysis of the plain text and comparison with the cipher text. The process is somewhat complicated, but not very difficult, and certainly useful, for it will solve all other messages in the same key. The process of anagramming messages of equal length is, of course, not restricted to this particular transposition cipher, but can be applied to all other methods of transposition in which messages of identical length are at hand.

69. Messages with similar beginnings.—It often happens that two messages will begin with exactly the same words, so that the first three or four lines of the enciphering rectangle are exactly the same in the two messages. When a single columnar transposition method using incompletely filled rectangles is being employed, the finding of two such messages will very greatly hasten the solution because the identical portions in the two messages enable the crypt-

analyst to divide up the cipher text into the exact columns of the enciphering rectangle, thus eliminating the doubts concerning the long and short columns that would otherwise hinder him in effecting a solution.

Note the identical portions in the two messages which follow:

Message 1—

[BNTSE] ARKC[L CET]TN BIT[ER RO]TAE LT[NNO N]NENO
O[TOKM] SZTGN [YITD]K LAN[AE FTFS]N PGNP[A RWO]IA
OFG[TF CT]OTD NI[NOE WX]ERF AS[IOS T]IDRR R[MMAO]
ARPAT [OUTI]O BIEO[A GAA]PN EIK

Message 2—

[BNTSE] INDOT [LCET]S AFPL[E RRO]MO ISOE[N NON]ST
IIU[TO KM]FEY KPC[YI TD]VSI NT[AEF TFS]TO NTN[AR
WO]ARO EEK[TF CT]TLT AEA[NO EWX]PV TIT[IO ST]TTF
OC[MMA O]OSCA NR[OUT I]EELS O[AGAA] ABITR T

FIGURE 56.

Let us now rewrite these messages so as to bring the identical portions all on the same lines. Note that the letter A of the identity AEFTFS is a mere coincidence, and does not belong to the identical portion. Such coincidences are common in this method of analysis and must be taken into consideration when "cutting" the text into the columns:

MESSAGE 1.														MESSAGE 2.													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14
B	L	E	N	T	Y	E	A	T	N	I	M	O	A	B	L	E	N	T	Y	E	A	T	N	I	M	O	A
N	C	R	N	O	I	F	R	F	O	O	M	U	G	N	C	R	N	O	I	F	R	F	O	O	M	U	G
T	E	R	O	K	T	T	W	C	E	S	A	T	A	T	E	R	O	K	T	T	W	C	E	S	A	T	A
S	T	O	N	M	D	F	O	T	W	T	O	I	A	S	T	O	N	M	D	F	O	T	W	T	O	I	A
E	T	T	N	S	K	S	I	O	X	I	A	O	P	E	S	M	S	F	V	S	A	T	X	T	O	E	A
A	N	A	E	Z	L	N	A	T	E	D	R	B	N	I	A	O	T	E	S	T	R	L	P	T	S	E	B
R	B	E	N	T	A	P	O	D	R	R	P	I	E	N	F	I	I	Y	I	O	O	T	V	F	C	L	I
K	I	L	O	G	N	G	F	N	F	R	A	E	I	D	P	S	I	K	N	N	E	A	T	O	A	S	T
C	T	T	O	N	A	N	G	I	A	R	T	O	K	O	L	O	U	P	T	T	E	E	I	C	N	O	R
						P		S						T	E		C	A	N	K	A	T		R		T	

FIGURE 57.

It is clear from a comparison of these diagrams, and a consideration of the fact that the long columns must of necessity go on the left hand side of the rectangle, that the numbers 7 and 10 occupy the first two positions in the key, and that the numbers 2, 4, 11, and 13 occupy the last four positions in the key. By anagramming columns 7, 10, and columns 2, 4, 11, and 13, we easily determine the exact order

of these numbers; by adding to the portions of words formed by this anagramming, solution may be completed. Thus:

MESSAGE 1.		MESSAGE 2.	
7-10	2-11-13-4	7-10	2-11-13-4
E N	L I O N	E N	L I O N
F O	C O U N	F O	C O U N
T E	E S T O	T E	E S T O
F W	T T I N	F W	T T I N
S X	T I O N	S X	S T E S
N E	N D B E	T P	A T E T
P R	B R I N	O V	F F L I
G F	I R E O	N T	P O S I
N A	T R O O	T I	L C O U
P S		N T	

FIGURE 58.

The completed rectangles are as follows:

MESSAGE 1.	MESSAGE 2.
7-10-3-12-6-11-4-9-5-8-2-11-13-4	7-10-3-12-6-11-4-9-5-8-2-11-13-4
ENEMYBATTALION	ENEMYBATTALION
FORMINGFORCOUN	FORMINGFORCOUN
TERATTACKWESTO	TERATTACKWESTO
FWOODSATMOTTIN	FWOODSATMOTTIN
SX TAKEPOSITION	SX MOVEATFASTES
NEARLANTZANDBE	TPOSSIBLERATET
PREPAREDTOBRIN	OVICINITYOFFLI
GFLANKINGFIREO	NTSANDTAKEPOSI
NATTACKINGTROO	TIONTOREPELCOU
PS	NTERATTACK

FIGURE 59.

70. Messages with similar endings.—The solution of two messages in which the plain-text endings are identical is even more simple. Here the bottom lines of the rectangle contain the same letters and afford clues to a direct reconstruction of the key. Note the identities in the two messages below:

Message 1—

¹⁼⁷ ETRTE E[ES]OA ²⁼⁶ AEUN[I V]AFLN ³⁼⁹ IA[MN]D RYHRV ⁴⁼² [ME]NRI
⁵⁼¹⁰ EE[TRO] ⁶⁼⁵ UDCCC O[HT]CY ⁷⁼¹¹ MRRE[A R]HITN ⁸⁼³ DE[YEN] RNERV
⁹⁼⁴ S[RB]EN ¹⁰⁼¹² IGSK[A I]LNRA ¹¹⁼¹³ NF[NA]D ¹²⁼¹ ALOLT ¹³⁼⁸ [XO]MAH HRR[EI]

Message 2—

¹ TLVS[X O]PNRE ² [ME]FDS ³ K[YEN]R ⁴ UEE[RB] ⁵ TSRE[H T]IANT
⁶ [IV]YMR ⁷ V[ES]IR ⁸ EEN[EI] ⁹ NOLT[M N]NEDE ¹⁰ [TRO]OP ¹¹ UN[AR]A
¹² CIA[AI] ¹³ NSCW[N A]

FIGURE 60.

The numbers above the brackets show the equivalency between the indicated portion in Message 1 and its identity in Message 2. The series of equivalents is as follows:

Message 1 = 1-2-3-4-5-6-7-8-9-10-11-12-13

Message 2 = 7-6-9-2-10-5-11-3-4-12-13-1-8

Now Message 1 has 105 letters; since the key consists of 13 numbers (indicated by the 13 identities), the rectangle for Message 1 contains 12 columns of 8 letters and 1 column of 9 letters. Message 2 has 81 letters, and its rectangle contains 10 columns of 6 letters and 3 columns of 7 letters. The rectangle of Message 1 has but 1 long column, whereas that of Message 2 has 3 long columns. Relative to the position the last letter in each rectangle occupies in the last line of the rectangle, it is obvious that the last letter of rectangle 2 is 2 letters in advance of the last letter of rectangle 1. Using this difference, viz, 2, let us build up a key sequence from the series of equivalents given above. Thus, the equivalent of identity 1 of Message 1 is identity 7 of Message 2, and we place the number 7 two intervals to the right of the number 1; the equivalent of identity 7 of Message 1 is identity 11 of Message 2, and we place the number 11 two intervals to the right of number 7, and so on until we obtain the following sequence:

1-2-3-4-5-6-7-8-9-10-11-12-13

1- 7- 11- 13- 8- 3- 9

The equivalent of identity 9 of Message 1 is identity 4 of Message 2, and the number 4 is placed between the numbers 1 and 7 in this sequence, for we can regard the sequence as being in the nature of a cycle or a continuous series. From this point on, the process is the same as before, and we finally have the following:

1-2-3-4-5-6-7-8-9-10-11-12-13

1-4-7-2-11-6-13-5-8-10-3-12-9

After little experiment it becomes obvious that column 8 belongs on the extreme left and that the key is 8-10-3-12-9-1-4-7-2-11-6-13-5. The completely deciphered messages are shown in Figure 61.

8-10-3-12-9-1-4-7-2-11-6-13-5

HEADREDCOLUMN

INFANTRYANDAR

TILLERYMARCHI

NGNORTHREACHE

DSILVERRUNCRE

EKATSEVENFORT

YAMXREMAINHER

EINOBSERVATIO

N

8-10-3-12-9-1-4-7-2-11-6-13-5

INFANTRYPOINT

REDCOLUMNPASS

EDSILVERRUNCR

EKATSEVENTWE

NTYAMXREMAINH

EREINOBSERVAT

ION

FIGURE 61.

The possibility of the rapid solution of such transposition ciphers by means of the method of similar beginnings and endings, as well as of identical lengths, constitutes one of the most serious drawbacks to the use of transposition ciphers in military cryptography, because it is almost impossible to avoid such cases where many messages must be sent in the same key each day.

71. Concluding remarks on transposition ciphers.—It was formerly thought that transposition ciphers are not suitable for military use because of the ease and rapidity with which they may be solved. But it is quite likely that certain transposition ciphers of the two-step type may be used more often in future operations, because their solution can be made extremely difficult without much loss in rapidity and simplicity in operation. However, it must be stated in qualification of the preceding statement, that such ciphers can never be used where the volume of traffic is large because of the great danger of messages of exactly the same length being intercepted by an alert enemy. If absolute supervision over all the messages in the same key could be exercised, the system might be practicable for heavy traffic, but the difficulties of exercising such a supervision in the field of operations are very great.

SECTION XVII.

CONCLUDING REMARKS ON MILITARY CIPHERS.

	Paragraph.
Combined substitution-transposition ciphers.....	73
Determining the cryptographic system.....	73
Synoptic table for cipher analysis.....	74
The requirements which a field cipher should fulfill.....	75

72. Combined substitution - transposition ciphers.—A method of substitution may be applied to a message, followed by the application of a method of transposition, or the reverse order of treatment may be employed. The resulting cryptogram would be extremely difficult to reduce if both methods were well chosen.

Combined substitution-transposition ciphers are, however, not often encountered in military cryptography. The errors ever present in the preparation and transmission of cryptograms, the time necessary for the operations, and the requirement of simplicity of method, so that the steps involved in encipherment and decipherment will not be beyond the ability of the average cipher clerk, all work to exclude the majority of combined methods. Where a combination is used, one or both of the methods must be simple if the system is to be at all practicable.

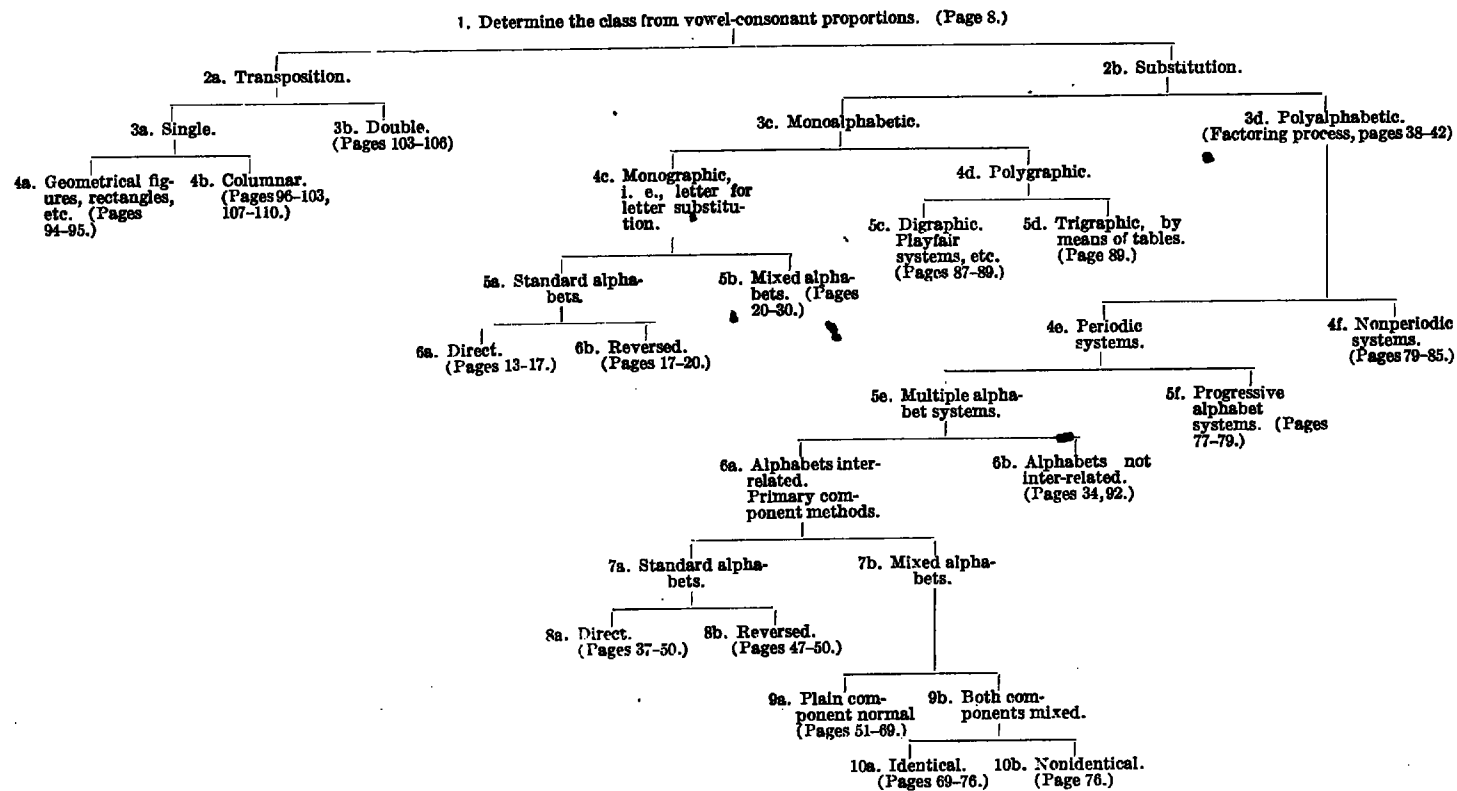
An example of such a system which was used with considerable success by the Germans in the late war is interesting. The message was first enciphered by means of a rectangle similar to that illustrated on page 33, using a square of 36 sections so as to include the alphabet and the 10 digits. The letters and figures were in a mixed order within the square, and the sides of the square were identified by the six letters, A, D, F, G, V, and X. Thus, bilateral substitution resulted. The substituted text was inscribed within a rectangle and columnar transposition by a numerical key applied. The length of the transposition key varied between 12 and 23 numbers, and the transposition key, as well as the arrangement of the letters in the substitution rectangle, was changed every two or three days. This cipher was used only for the more important tactical messages between the larger headquarters, usually between division, corps, and army.

This method results in the production of a cryptogram in which the text consists of the *transposed halves of substituted pairs* for the letters of the plain text. Solution requires two steps: First, the transposition text must be arranged so as to bring the proper halves of letters together, and then the substitution text must be analyzed on the basis of frequency and repetitions. The methods employed in the analysis are extremely interesting but too long to find a place in this book. It may be added that more than 70 per cent of the messages sent in this cipher were solved, yielding information of great value.

73. Determining the cryptographic system.—It is obvious that the first step in the analysis of a cipher is to determine the general system of encipherment. Most of the ciphers produced by the simple systems described in the foregoing pages show, either externally or internally, definite characteristics by means of which the cryptanalyst can determine what basic system of encipherment was employed. But in the field of complex ciphers, this determination is the most difficult part of the analysis, and, unfortunately, there are no absolutely definite tests that can be applied to the cipher text which will disclose the system. In the case of military cryptography, the system sooner or later becomes known to the enemy, either by capture of alphabets, tables, plain-text messages, apparatus, prisoners, etc. But in the case of isolated messages, where the system is of a complicated, nonperiodic type, the only method of determining the system is often by the long and laborious process of elimination. The science of cryptanalysis has not as yet reached the stage of refinement of methods such as are employed, for example, in qualitative analysis in chemistry, and this is what makes the analysis of cryptograms in unknown complex systems so difficult. Without a knowledge of, or a shrewd guess as to the system

employed, the cryptanalyst is unable to decompose the heterogeneous text of the cryptogram into the homogeneous distributions of single frequency tables, and this, we have seen, is the ultimate and absolutely necessary step in the analysis. The only knowledge that the cryptanalyst can bring to his aid in this most difficult step is that gained by long experience and practice in the analysis of many different types of systems.

74. Synoptic table for cipher analysis.—The accompanying Table 5 is a diagrammatic résumé of the various types of ciphers treated in this pamphlet. It may be of assistance to the student in his earlier work in cryptanalysis, in that it shows in a compact form the general relationship existing between the few ciphers discussed herein, and refers him specifically to the pages on which the principles and method of their analysis are presented. It is admittedly a very brief outline, and can therefore be of but little assistance in the analysis of the more complex types of ciphers he may encounter in more advanced work. For expositions of certain of the more difficult types of ciphers the bibliography given at the end of this book should be consulted.

TABLE 5.—*Synoptic table for cipher analysis.*

75. The requirements which a field cipher should fulfill.—The student is now in a better position to understand the reasons underlying the requirements which a cipher that is to be used in the field of operations must fulfill than he would be if these requirements had been stated in the opening pages of this book. The principal requirements are given below, and it should be added that there are also some others which must be taken into consideration.

(a) The cipher message must be in a form suitable for telegraphic transmission. This involves two features:

- (1) It eliminates all forms except those composed exclusively of either letters or figures.¹ These, as a rule, are transmitted in groups of five.
- (2) It eliminates all those systems in which the cipher text is longer than the plain text. This is obviously necessary, for purposes of economy and rapidity.

(b) It must be granted that the enemy is in possession of all the details of the general method of encipherment; that is, the general system is known to him. The only thing kept secret is the specific key applying to the messages. This key must therefore be of such a nature that it can be easily varied, and it may consist of a letter, word, phrase, sentence, or number easily remembered or derived.

(c) The method of encipherment must not be complicated nor require the application of a long series of rules, and must be such as to require the least possible mental strain on the part of the operator. As a rule, complex double processes are not suited to the conditions in the field of operations, but occasionally systems involving not more than two steps, if each be simple and rapid, possess advantages. If a piece of apparatus or a machine is used, it must be small enough to be carried about without inconvenience.

(d) The system must be such that the errors, which inevitably occur in cryptographic communication, can be corrected easily and rapidly. This requirement will generally rule out all systems in which an error or the addition or omission of a single letter will affect several letters.

(e) Conforming to all the preceding requirements and assuming that the enemy will be in a position to intercept 75 per cent of each day's traffic by radio, which may consist of 50 to 200 messages *all in the same key*, nevertheless the messages should still be proof against unauthorized decipherment for such a length of time as will render the information thus obtained of no value.

¹ That is, the message, if composed of letters, must not contain any figures, punctuation signs, or other symbols whatsoever; if composed of figures, it must not contain any letters, punctuation signs, or other symbols. The cost of transmitting messages composed of intermixtures of letters, figures, etc., is prohibitive.

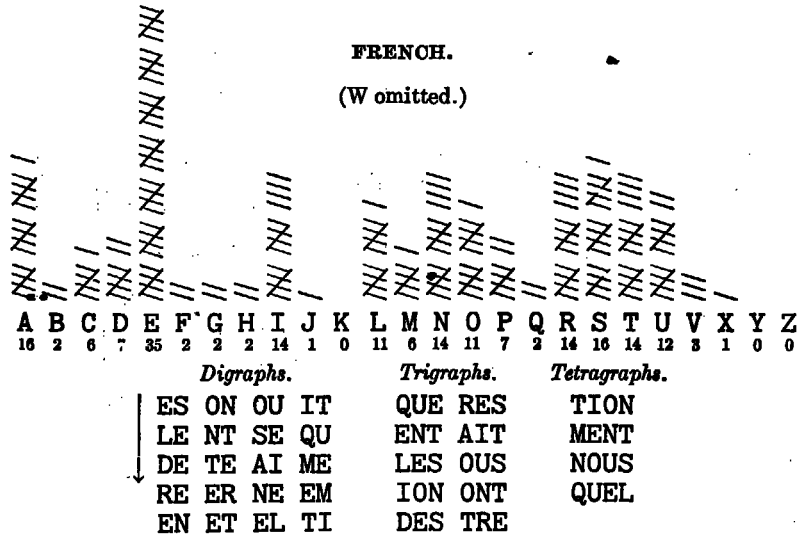
SECTION XVIII.

FREQUENCY DATA FOR OTHER LANGUAGES.

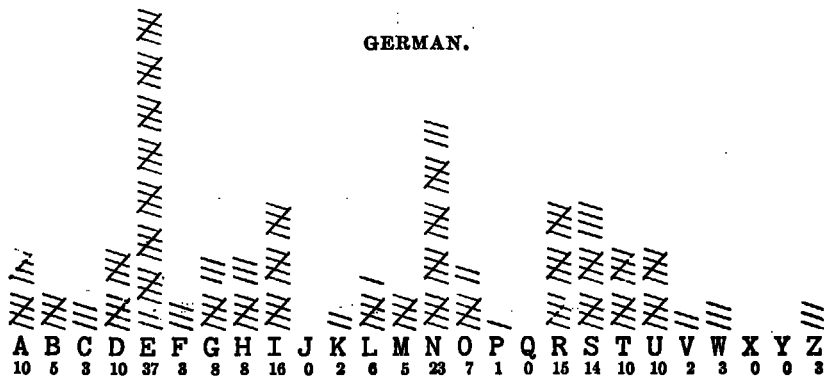
Basis of tables.....	Paragraph.
French.....	76
German.....	77
Spanish.....	78
Italian.....	79
Portuguese.....	80
Japanese.....	81
Russian.....	82
	83

76. Basis of tables.—The following frequency data for the principal foreign languages, French, German, Spanish, Italian, Portuguese, Japanese, and Russian, are given for reference. The order of the letters in all these alphabets, except the last two, is the same as in our own alphabet, but certain letters are omitted, or exceedingly rare, and are found only in proper names and borrowed words. All the frequencies given are on the basis of 200 letters, the proportions having been reduced to this basis from a much larger count of letters.

77. French.



78. German.



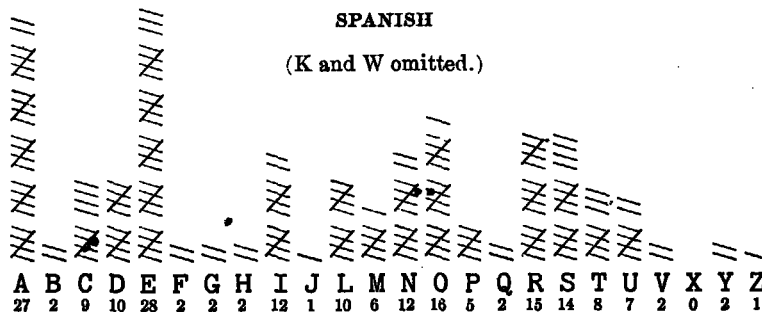
ELEMENTS OF CRYPTANALYSIS.

<i>Digraphs.</i>	<i>Trigraphs.</i>	<i>Tetragraphs.</i>
EN EI EL ND	EIN SCH NEN	ICHT CHEN
ER IE TE UE	ICH CHE DES	KEIT CHER
CH IN UN SE	DEN DIE BEN	HEIT URCH
DE NE ST AU	DER UNG RCH	CHON EICH
GE BE DI RE	TEN GEN	
	HE	
	CHT UND	

The unlauded vowels of ordinary text, ä, ö, and ü, are usually replaced in telegraphic and cryptographic messages by their equivalent diphthong combinations with the letters e, i, or u.

79. Spanish.

SPANISH
(K and W omitted.)



A	B	C	D	E	F	G	H	I	J	L	M	N	O	P	Q	R	S	T	U	V	X	Y	Z
27	2	9	10	28	2	2	2	12	1	10	6	12	16	5	2	15	14	8	7	2	0	2	1

Digraphs. (Valerio.) *Trigraphs.*

ES	AR	ON	SE	QUE	DEL	PER
EN	UE	QU	TA	EST	CIO	IST
EL	RA	ST	CO	ARA	NTE	NEI
DE	RE	AD	CI	ADO	OSA	RES
LA	ER	AL	IO	AQU	EDE	SDE
OS	AS	OR	NO			

80. Italian.

ITALIAN.

(J, K, W, X, and Y omitted.)

i
c
o
a
r
-
s
s
m

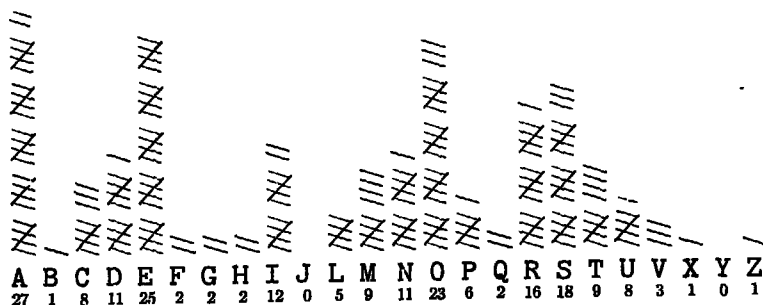


A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T	U	V	Z
23	1	9	6	25	2	3	1	23	14	6	13	19	6	1	13	10	12	6	4	2

<i>Digraphs.</i>	<i>(Valerio.)</i>	<i>Trigraphs.</i>
ER DE AN LE		CHE QUE ESI
ES DI RA TO		ERE ARI IDI
ON TI NT IO		ZIO ATO ERO
RE SI TA AR		DEL EDI PAR
EL LA CO NE		ECO IDE NTE
EN AL IN OR		

81. Portuguese.—The following data were extracted from an original report by Maj. O. Holstein, O. R. C.

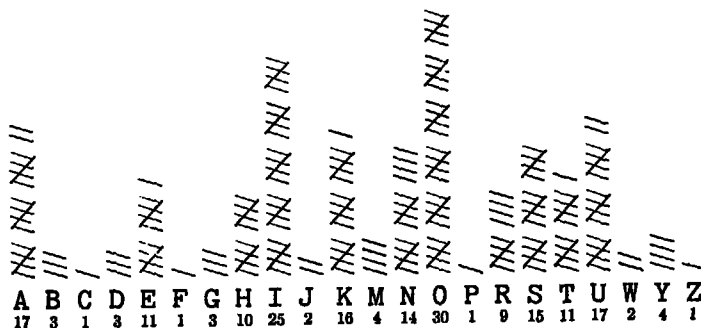
PORTUGUESE.
(K and W omitted.)



<i>Digraphs.</i>	<i>Trigraphs.</i>
ES CO AD AR	QUE DES ADE
OS DO NT TE	ENT ODE ARA
DE EN SE EM	NTE STA COM
RO RE OR QU	ADO CON RES
AS ER AO SA	EST MEN DOS

82. Japanese.—The frequencies of the Roman letters used in expressing the so-called "J kana" form of the Japanese language are given below.

JAPANESE.
(L, Q, V, and X omitted.)



83. Russian.—The symbols employed in the Russian alphabet are somewhat different from those of the other foreign languages. This alphabet consists of 36 letters, but the special adaptation of the Morse telegraph code to this alphabet contains but 31 letters whose frequencies are as follows:

Frequencies.	Russian.	Equivalent.	Morse.
///	А	A	..-
////	Б	B	.-...
///	В	V	..---
////	Г	G	---.
///	Д	D	---.
///	ЕЭ	E	..
///	Ж	J	...-
///	З	Z	---..
///	ИИ	I	..
///	Й	I	..---
///	К	K	..-
///	Л	L
///	М	M	---
///	Н	N	---
///	О	O	---
///	П	P	..---
///	Р	R	..
///	С	S	...
///	Т	T	---
///	У	U	..-
///	ФФ	F
///	Х	KH
///	Ц	TS	..---
///	Ч	CH	---
///	Ш	SH	---
///	Щ	SHCH	---
///	ЪЪ	MUTE	---
///	Ы	I	..---
///	Ь	YE
///	Ю	YU	..---
///	Я	YA	..---

Digraphs.

НО СТ ЕН ОС ОБ ПО РА НА ТО ГО ПР НЕ ВО НИ КО

Trigraphs.

СТВ ЕНІ ЕНН ОСТ ПОЛ ЕЛЪ СТА ТЕЛ ННО ЛЬН

PART 2.—THE ANALYSIS OF CODES.

SECTION XIX.

CODE SYSTEMS AND CODE BOOKS.

	Paragraph.
Code systems in general.....	84
Operations of encoding and decoding.....	85
Types of code books.....	86
Types of code groups.....	87
Permutation tables and the "two-letter differential".....	88
"One-part" and "two-part" codes.....	89
Purposes of the two-part type of code.....	90
Enciphered code.....	91

84. Code systems in general.—We come now to a discussion of the nature, uses, and types of code systems used in communication. Considered in its broadest aspect, a code system is really only a specialized type of substitution cipher, and, in fact, there are some systems of code which so closely approach cipher systems that no sharp line of demarcation can be established to separate code from cipher systems. As was explained in Section I, the essential difference between the two systems lies in the fact that in cipher systems we deal with units of equal length (single letters, pairs, or groups of definite length), applying some form of transposition, or substitution, or a combination of the two principles to these units. In code systems we deal with units of unequal length (ranging from single letters to entire sentences), arbitrary, equal length combinations of letters or figures provided by a code book being substituted for those units. The one system may necessitate the use of no apparatus whatsoever, other than pencil and paper; the other system requires the possession of identical copies of a *code book* by all the correspondents. Code books are often simply termed *codes*.

85. Operations of encoding and decoding.—The simplicity of code as a system of communication is one of its chief advantages. In encoding it is necessary merely to replace the various words, phrases, numbers, etc., by the letter or figure groups as provided by the code book. In the case of words or names which are not already in the vocabulary, provision is made for building up the word by means of syllables and individual letters. It is usually the case that the encoded message is somewhat shorter than the original plain-text message on account of the abbreviating nature of code; sometimes a single code group will represent a long phrase of perhaps five times as many letters. This feature, of course, constitutes one of the most important advantages of code from a commercial point of view.

The process of decoding is, of course, merely the reverse of encoding, and, where the errors in transmission are few, is very rapid. It is obvious, however, that even a small number of errors in a message may obscure the meaning, or render it extremely difficult to decode.

86. Types of code books.—There are various types of code books, depending upon their uses. We are all more or less familiar with the ordinary kinds of commercial and business codes, used extensively for the purposes of economy, such as the ABC Code, Lieber's Code, Bentley's Code, Western Union Code, and the like. They are usually fairly large codes adapted for general commercial correspondence. Most large business firms have their private codes, constructed especially for their use, and containing a more or less highly specialized vocabulary. If its circulation is very limited such a code may also constitute a secret code. There are also many commercial codes which are adapted to a particular industry, for example, the rubber or the sugar industry, and can be purchased by the general public from the publishers. Such codes usually have a highly specialized technical vocabulary in addition to the general vocabulary.

We are, however, more concerned here with such codes as are employed in military or naval communication, and while the general resemblance between the ordinary commercial codes and the usual governmental codes is considerable, yet the primary purposes are different in the two cases. The principal purpose of code in commercial practice is to effect economy in transmission, secrecy being usually of secondary importance; its principal purpose in governmental affairs is to effect secrecy, and economy, while an additional desirable feature, is of secondary importance. These differences in purpose occasion the introduction of certain features in one type that are not present in the other. This will be discussed under paragraph 90.

87. Types of code groups.—As regards the types of code groups used in codes, they are of two general classes: (a) Letter groups, (b) figure groups. Both possess advantages and disadvantages. In those parts of the world where italic or Roman letters are used for writing, letters possess greater advantages as regards accuracy in reading by telegraph operators, this being the prerequisite to correct transmission and reception. However, in some parts of the world, for example, Turkey and Russia, telegraph operators, except in the larger cities, are unfamiliar with our italic or Roman letters, and hence many errors arise. But the Arabic digits are almost universally recognized and used, so that for communications between obscure ports and cities in certain foreign countries, figure groups are preferred over letter groups. Many code books, however, contain both figure groups and letter groups, so that either may be used at the discretion of the correspondents.

The greatest advantage possessed by letter groups over figure groups, however, lies in the availability of a far greater number of permutations of letter groups, because there are 26 letters which may be permuted to form letter code groups, whereas there are only 10 digits which may be permuted to form figure code groups. If code groups of five elements are used, then there are available 26^5 , or 11,881,376 groups of five letters, and only 10^5 , or 100,000 groups of five figures. Now since the number of permutations of 26 letters taken in groups of five is so great, only permutations conforming to special types may be selected for use, and there will still remain a sufficient number of code groups for even the largest codes. The selection of certain types of code groups is done with a view to reducing to a minimum the inevitable errors in telegraphic transmission. Furthermore, if the code groups have been constructed scientifically it is possible to provide a quick and effective means of correcting such errors as do creep in without having to call for a repetition of the message.

The length of code groups used, i. e., whether they are groups consisting of two, three, four, or five letters, or figures, depends upon the size of the code. This, however, applies almost exclusively to field military or naval codes, where transmission is through a governmental agency; for in commercial messages or in governmental communications transmitted over privately owned and operated lines, five-letter or five-figure groups are used almost exclusively on account of the regulations adopted by the commercial telegraph and cable companies. For cable communications it is permissible to join two five-letter groups, which will be charged for as one word, provided the resulting group of ten letters is pronounceable according to the common usage in any one of the following languages: English, French, Spanish, Italian, German, Portuguese, Dutch, or Latin. If the group is not pronounceable then each group of five letters counts as one word, whereupon it is seen that the cost of cablegrams using non-pronounceable groups is double what it might be. All governments, and all commercial firms take advantage of this regulation wherever possible. In ordinary land line transmission of either code or cipher messages, each group of five letters, whether pronounceable or not, is charged for as one word.

88. Permutation tables and the "two-letter differential."—Now it would be very difficult to find a large number of intelligible words all of five letters, in the various languages which are permissible. Therefore, in modern codes the code groups are merely arbitrary, pronounceable, five-letter groups constructed by the use of PERMUTATION TABLES, which permit of the systematic construction of code groups of the type desired. Inasmuch as those codes which are based upon permutation tables show the table and explain how to use it, it is unnecessary to go further into this subject.

Such code groups, when scientifically constructed, have an additional feature, and one which has greatly increased the reliability of code as a system of communication. To make one error in a group of five letters is not at all unusual on the part of the average telegraph operator. If only a difference in one letter distinguishes one code group from another in a given code book, e. g., ABABA and ABABE, then serious errors may be introduced in the meaning of a message, or the message may be rendered unintelligible or obscure by the occurrence of only a few telegraphic errors. If, however, every code group in the book is distinguished from every other code group by a difference in at least two letters, then a telegraph operator would have to make two errors in a single group in order that a wrong meaning be conveyed by the mutilated group. This principle of having the code groups differ from each other by at least two letters is called the "two-letter differential."

89. "One-part" and "two-part" codes.—As regards their construction or arrangement, codes may be of two types:

1. "One-part," or alphabetical codes, in which the plain-text groups are arranged in alphabetical order accompanied by their code groups, which are also arranged in alphabetical order, or numerical order. Such a code serves for decoding as well as for encoding.

2. "Two-part," or randomized codes, in which the plain-text groups are arranged in alphabetical order, accompanied by their code groups arranged in a non-alphabetical or random order, the code groups being assigned to the plain-text groups in an absolutely arbitrary and random manner, by drawing the code groups out of a box in which they have been thoroughly mixed up, or by some other manner in which the element of chance operates in assigning the code groups to the plain-text groups. It follows, therefore, that such a list can serve only for encoding, and that for decoding, another list must be provided, in which the code groups are arranged in alphabetical or numerical order, accompanied by their meanings as given in the encoding section.

The following brief extracts from typical one-part and two-part codes will serve to illustrate the difference between them:

<i>One-part code.</i>	<i>Two-part code.</i>
	Encoding. Decoding.
ABABD A	GAJVY A ABABD Obstructed
ABACF Aaft	TOGTY Aaft ABACF Term
ABAHK Abandon	FEHIL Abandon ABAHK Branch
ABAJLit	BAYLTit ABAJL If it has not
ABALN Abandoned	WITYH Abandoned ABALN To be sent by
ABAMPby	NYSYXby ABAMP Acceding
ABAWZ Abandoning	IFWUZ Abandoning ABAWZ Building
ABBAD Abandonment	RUMGO Abandonment ABBAD Do not attempt

90. Purposes of the two-part type of code.—The two-part code is a comparatively recent development in code systems. Its purposes are twofold: (a) Greater secrecy, and (b) greater accuracy. These two features will now be explained.

In a one-part code the plain-text groups progress from A to Z in a regular alphabetical sequence, accompanied by their code groups, also in a regular alphabetical or numerical sequence. Thus, if the word ABAPT is represented by a code group whose initial letter is A, or whose initial number is 1, then the word ABANDON will be represented by a group whose initial letter is also A, or whose initial number is also 1. In other words, the enemy cryptanalysts have definite clues to follow in breaking down the code as a direct result of the two sequences progressing simultaneously; the determination of the value of one code group affords definite clues to the value of many other code groups. In a two-part code, however, the word ABAPT might be represented by a group whose initial letter is T, or whose initial number is 8, and the word ABANDON might be represented by a code group whose initial letter is F, or whose initial number is 3. In other words, the determination of the value of one code group affords no clues to the value of any other group, because the two sets of groups do not correspond in their progression from the beginning to the end of the code.

With regard to the greater accuracy of a two-part code over a one-part code, consider the following set of phrases which appear in a hypothetical one-part code:

WOVAM will be ready to attack.

WOVEN will not be ready to attack.

Such an arrangement is subject to two sources of error. A code clerk working under great difficulties, and in a hurry, may accidentally write down WOVAM instead of WOVEN, as a result of the contiguity of the two sets of letters which are nearly similar in appearance and are so close together on the page that his eye may take the group from the wrong line; or on account of the similarity in sound, his ear may deceive him into writing WOVEN when he should have written WOVAM. Now the meaning of the one group is the exact opposite of the meaning of the other and, since either meaning may fit in correctly with the context of the message, the error may remain undiscovered for some time, thus causing serious inconvenience or, in the case of combat, actual loss of life. Again, while the making of two errors in a single group is rather unusual in transmission or reception, yet it does happen and, in such a case as the above, the error would not be detected. But in a two-part code such errors are impossible. In the case of the first source of error mentioned above, the code clerk would be very much less likely

to confuse two entirely different groups of letters; in the second case, if two errors are made in the transmission or reception and if these errors involve the last two letters, producing a group which actually has a meaning in the code, this meaning is so unlikely to be such as to fit in correctly with the context that its probability of occurrence may be altogether neglected. Thus, if this sort of error does happen, the meaning of the group fails to fit in with the context and at once indicates that an error is involved. Knowledge of the existence of such an error, even if it is impossible to correct, is, of course, a much more preferable condition than is ignorance of its existence, with a possible action based upon an erroneous decodement.

Two-part codes are used by large governments for their secret diplomatic, military, and naval communications for the reasons given above, although the cost of compiling such codes is more than three times that of compiling the ordinary one-part codes.

91. Enciphered code.—Sometimes the code groups of a code message undergo a further process of encipherment, in which case the resulting cryptogram constitutes an *enciphered code message*. There are two circumstances in which enciphered code is employed. First, if the code book is not secret and it is desirable to transmit a secret message in this code, it becomes necessary to encipher the code groups. Secondly, even if the code book is kept secret, it is desirable, in the case of highly secret communications, to encipher the messages, in order to increase the degree of security by delaying as long as possible the reduction of the code by the enemy cryptanalysts.

In those codes which employ pronounceable groups for the sake of economy, it is obvious that the resulting enciphered code must remain pronounceable in order to take advantage of the reduced cost of cable messages, as previously explained. Hence, the systems of enciphering to produce this result are rather limited. But where the resultant form of the code groups is of no importance almost any type of encipherment may be applied. The augmented degree of secrecy due to the encipherment depends entirely upon the nature of the system applied.

SECTION XX.

ANALYSIS OF CODES.

	Paragraph.
Introductory remarks.....	92
Distinguishing between code and cipher messages.....	93
Factors in the analysis of codes.....	94
Steps in analysis.....	95

92. Introductory remarks.—Although the simplicity of code as a system of secret communication is one of its principal recommendations, there are, however, certain precautions which must be observed if the code is to be kept secret for any length of time. It might be

thought by the uninitiated that code messages, built up as they are by the substitution of purely arbitrary and meaningless groups of letters or figures taken out of a book which is kept secret, are wholly impossible to read by anyone not in possession of the code book. Such is by no means the case, as has been intimated in the preceding sections.

93. Distinguishing between code and cipher messages.—The student may first inquire concerning the method of distinguishing a code message from a cipher message. In only a few cases is the classification difficult, and these are only in the instances where single messages are intercepted. When many messages are available, all doubts are soon set aside. In the first place, when code books are used for regular service they are commonly of the type which employs pronounceable groups, as explained before, or groups of definite construction as regards the arrangement of vowels and consonants. There are, it is true, systems of cipher which produce pronounceable groups by employing a two-letter vowel-consonant or a consonant-vowel substitution for the letters of the plain text, but they are unusual. Such systems are sometimes called "pseudo code" systems, and closer investigation of the messages produced by them will disclose the fact that they are in reality cipher messages. The nature of this investigation is our next point.

Repetitions of code groups in a five-letter code of considerable size are rather rare in a single message or even in a few messages. When a repetition does occur, it is more likely to involve the whole group. The repetitions of letters in cipher messages are, of course, often many in number and always of varying lengths, and this forms the principal method of distinguishing a cipher from a code message in those rare cases in which the nature of the groups of the cryptogram does not disclose whether the cryptogram is a code or a cipher message. What has been said of code messages in a five-letter code applies also to those codes in which groups of less than five letters are employed, as in the case of certain field codes used in military or naval operations. In these cases, in fact, the interception of messages consisting of regular groups of two, three, or four letters is in itself an almost certain indication of code, since cipher is practically universally transmitted in five-letter or five-figure groups.

Cipher messages consisting of figures are not nearly so common as cipher messages consisting of letters, and when they are encountered the cryptographic system is apt to be of a more or less simple nature, resulting in the production of a considerable number of repetitions. Since the nature of such repetitions is similar to those encountered in letter ciphers, i. e., they are of variable length, what has been said in the preceding paragraph concerning the distinguish-

ing marks between letter codes and ciphers also applies in the case of figure codes and ciphers.

94. Factors in the analysis of codes.—Nearly all codes can be solved, the difficulty in analysis being dependent upon a number of factors, chief of which are the following:

(a) *The nature of the code.*—A purely alphabetical, or one-part code, is much easier to solve than a two-part, or randomized code. In the latter a "change" in code means merely the reassignment of code equivalents; the plain-text groups remain the same in nature, as well as in arrangement; the code groups remain the same also, but they are shifted about so that the plain-text groups are assigned entirely new code groups. Of course, with each new edition of the code, the decoding section must be changed to conform with the new values of the code groups in the encoding section. In a purely alphabetical code it is not possible to produce an entirely new edition from time to time to replace a compromised edition. At most one can only insert or eliminate a few words, but the sequence of plain-text groups and the sequence of code groups remain in general the same as before. The only thing that can be done in this case is to change the method of encipherment, or to introduce new enciphering tables from time to time. The solution of code messages in such a case becomes merely a slightly modified problem in cipher analysis, and may be achieved more or less readily according to the circumstances.

(b) *The size of the code.*—A code of only two thousand groups is much easier to solve than one of fifty thousand groups.

(c) *The volume of traffic available for study.*—If many messages are sent, a considerable amount of text will accumulate. Volume is dependent upon the length of time the code is in effect, and upon the average number and lengths of messages sent each day. The length of time during which a given code is in effect, or the "life of the code," varies usually with the size of the code book, and the amount of traffic. A large code used for a rather limited amount of traffic may continue in effect for a year or even several years. A small code used for considerable traffic may be unsafe to use for more than one or two weeks.

(d) *The intelligence system of the office or persons endeavoring to solve the code.*—The amount of information obtained through various sources and having a bearing upon the contents of the code messages is highly important in that such data form the basis for the assumption of values for many code groups. For example, in solving code messages sent between the various commands in the actual zone of operations, the location and the identity of the sending and receiving stations serving those commands may offer clues to the names and numerical designations of the units involved, and these will often facilitate the solution of important code groups, and lead to the identification of additional units mentioned in other code messages.

95. Steps in analysis.—The commonly employed expression “to break into” or “break down” a code is really not the best way of indicating what is involved in the analysis of code. The procedure is, in fact, a combination of an analysis, or “breaking down” process, and a synthesis, or “building up” process, leading to a *reconstruction* of the code book. However, for our purposes we shall retain the word analysis, bearing in mind that the opposite process of synthesis is also involved. The ultimate solution is usually accomplished as a result of the study of a comparatively large volume of text obtained through radio intercepts, or other means. In the early stages of the process, the analysis enables the resolution of the code text into its principal elements, with an accompanying partial reconstruction of the code book; some messages may be read in part, others in whole. As the work progresses, however, and unless a new edition of the code comes into effect, the reconstruction is more and more nearly completed until finally the cryptanalyst is in a position to read every one of the messages as easily and as quickly as are the legitimate or intended recipients. Such a reconstruction may be accomplished by two different methods, the second, however, being dependent upon previous success by the first method. They are—

1. Solution by detailed analysis, or “First Principles.”
2. Solution by comparison or analogy.

In what follows we shall treat more specifically of the tactical field codes in use in the zone of operations rather than the larger codes used for administrative correspondence, although the fundamental principles are the same in both cases.

SECTION XXI.

SOLUTION BY DETAILED ANALYSIS OR FIRST PRINCIPLES.

	Paragraph.
Use of repetitions in codes as compared with their use in ciphers.....	96
Compiling and studying the data.....	97
Process of classification.....	98
Process of identification.....	99
Comparison of solution of a one-part code with a two-part code.....	100
General remarks on detailed analysis by first principles.....	101

96. Use of repetitions in codes as compared with their use in ciphers.—We have seen that the solution of most ciphers is attained by a study of frequency tables which show the repetitions of single letters, digraphs, and trigraphs. The solution of code messages is likewise attained by an understanding of the principles of frequency, not only of single letters or of combinations of letters, that is, syllables, but also of words and combinations of words, that is, phrases. Just as certain individual letters or certain digraphs and trigraphs are used more frequently than others in ordinary intelligible text, so certain words and phrases are used more frequently than others.

The various parts of speech of which sentences are composed, the nouns, verbs, adjectives, prepositions, etc., all have their characteristic frequencies, which may be established or determined by actual counts. Just as the letter E is usually the most frequently occurring letter in normal English text, so the word THE is the most frequently occurring word in normal English text. Again, certain letters show distinct peculiarities; the letter H, for example, is most often preceded by the letter T; the letter Q is nearly always followed by the letter U, etc. Likewise, we find certain peculiarities in phrase formation; the word THE is most often preceded by a preposition such as IN, OF, or BY; the word TO is often followed by a verb, etc. Moreover, while single letters may combine in a multitude of ways to form words, in intelligible phrases and sentences the possible combinations of words are rather limited. In other words, it is found that certain groups of words tend to form more or less set or unvarying sequences. For example, the letters A, R, and E may form words or parts of words in six possible permutations of these three letters: ARE, AER, REA, RAE, EAR, ERA. But the words REFERRING, TO, and YOUR can form an intelligible phrase in only one combination: REFERRING TO YOUR. Again, just as certain letters are used more frequently than other letters to begin words, so certain words are used more frequently than others to begin sentences. Furthermore, although a word may begin or end with almost any one of the 26 letters of the alphabet, a sentence can not begin or end with just any words at all. Questions usually begin with an interrogative, such as HOW, WHEN, WHERE, or with certain verbs, such as IS, ARE, CAN, WILL, etc. In short, all the manifold peculiarities of phrase and sentence formation are seized upon and made use of by the code solver.

97. Compiling and studying the data.—But the most important feature of the analysis and reconstruction is the necessity of concentrated, persistent study of voluminous records, indexes, charts and notes. The methods of recording and tabulating the data are very important and depend upon the nature of the code. It is essential to have the most important data in a very condensed and concentrated form, so that the eye may be able to survey in one glance a great array of facts and characteristics of the text. The most insignificant details may lead to far-reaching clues, and an exceedingly trifling error on the part of a single code clerk may be the entering wedge to a quick solution. Upon the methods of marshaling together the data, the skill and experience of the expert, his knowledge of the language, blunders of the enemy, and finally, luck, will depend the success and speed in the solution. There is not the space in this brief treatise to give in detail the methods of analysis. We may indicate them only in outline.

The reconstruction of the code book by this laborious process goes through two more or less distinct stages:

1. Classification of code groups into certain categories.
2. Identification of the members in each category.

98. Process of classification.—When the code analyst first begins his study of the text, he is confronted with a body of completely unintelligible sentences to which he must apply certain tests more or less equivalent to the reagents in chemical analysis. By means of these tests he is able to break up the text into definite classes or sets of groups. The code groups of most codes may be classified into five or more distinct types, based upon their behavior in the code messages. They may be the equivalents of—

- (a) Numbers.
- (b) Words, phrases, sentences.
- (c) Spelling groups, that is, single letters or syllables.
- (d) Punctuation.
- (e) Auxiliary signals, such as indicators of various sorts.
- (f) Nulls, or nonsignificant groups, inserted to throw the enemy solvers off the track.

By a careful study of the behavior of frequently recurring code groups, the expert is led to conclude that certain groups, because of the general characteristics they exhibit, must be the representatives of numbers, others of spelling groups, others of punctuation, and so on. This, then, is what is meant by the process of classification. When classification has proceeded upon a solid foundation far enough, each set of groups is underlined throughout the text in some distinctive manner by means of colored pencils. This has the effect of setting forth very clearly to the eye the various parts of speech of which each message is composed.

99. Process of identification.—Subsequent to this, the individual members of each class of code groups are subjected to closer scrutiny. Classification is based upon a knowledge of the *general behavior* of the various classes of groups; code groups representing numbers behave differently than do those which represent punctuation, for example. Identification is based upon a knowledge of the *specific behavior* of the various elements in each group; the word AND for example, behaves very differently from the word PERIOD or the word IS. It will be impossible to set forth specific instances, but suffice it to say at this point that after an intensive study along these lines one is able to make assumptions for the values of certain code groups in each set. If the assumed values, when applied throughout all the messages wherever the groups in question occur, yield intelligible text, one has at last attained the solution of a few code groups. Once a few correct values have been established, new values are soon obtained from the positions in which the identified groups are

found in other messages, and the process goes on in a cumulative manner, one solution leading to the solution of several other groups until many identifications have been established.

100. Comparison of solution of a one-part code with a two-part code.—This process of analysis is much easier in the case of the one-part code than in the case of the two-part code, for reasons already stated. In fact, the solution of a one-part code where the amount of text available for study is considerable may be easier than the solution of the more complex types of substitution ciphers. In the first place, once a few messages have been intercepted and the lower and upper limits of the code groups ascertained, the cryptanalyst may automatically block out sections of groups to correspond with the frequency of letters of the alphabet with regard to their use as the initial letters of words. For example, in English, words beginning with the letter A are much more numerous than words beginning with the letter B, and the latter in turn are more numerous than those beginning with the letter J or Q, and so on. The range of letter groups or figure groups of which the messages are composed may thus be split up into 26 sections proportionate to the number of words with the initial letters A, B, C, and so on. A group such as AKAZO may therefore be assumed to be the equivalent of a word beginning with the letters AC, whereas another group such as ZYTIP may be assumed to be the equivalent of a word beginning with the letters YO. These, of course, are but rough examples to illustrate the point. Codes of the two-part type can not, however, be blocked out in this manner and are much more difficult to reconstruct on this account. In the recent war the Germans employed both types of codes, a small one-part code for front-line work and a larger two-part code for field operations up to division. The one-part code was enciphered by a rather simple system, so easy to break down that the enciphered code messages of this code were usually solved within a few hours after three or four messages enciphered by the new enciphering table had been intercepted. On the other hand, it took from ten to twenty days to reconstruct a sufficient amount of the two-part code to enable solutions of any value to be achieved.

101. General remarks on detailed analysis by first principles.—Once a code has been solved in this manner, after a long and painstaking investigation, the knowledge of the general contents and more particularly the general form of the messages is of the greatest assistance in the solution of subsequent editions of the code (of the two-part type). It is for this reason that the adoption of a more or less set or stereotyped form of messages and phraseology is a very dangerous procedure from the viewpoint of maintaining the secrecy of the code. The formation of habits and characteristic

forms of expression on the part of code clerks in the preparation of code messages is, therefore, of the greatest assistance to the enemy. A tendency to the promiscuous use of punctuation, such as PERIOD and COMMA is one of the greatest faults. It may be thought that the use of punctuation in code messages is of no assistance to the enemy because in themselves they possess no significance, but such is far from the case. Instead of leaving the identification of these groups to the very end of the process of solution, as may be imagined at first hand by the uninitiated, on the supposition that their solution is merely one of the final refinements and unnecessary additions to the meaning of messages, the cryptanalyst regards the solution of the code groups representing punctuation as one of the most important preliminary steps to the solution of the text, because of the exceedingly valuable clues it affords in the classification and identification of the code groups representing words, numbers, or spelling groups. Therefore, all punctuation not absolutely necessary to the intelligibility of the message should be omitted.

In this connection it may be stated that so far as the cryptanalyst is concerned, *all messages are of equal importance*, for it is often the case that messages of the least significance in the tactical situation lead to the greatest and most far-reaching clues in analysis. For example, the practice messages sent by the German operators furnished a great part of the data for the reconstruction of their codes, and these messages consisted largely of proverbs, greetings, jokes, and the like. It is amusing to record that the German proverb "Morgen Stunde hat Gold in Munde," the equivalent of the "The early bird catches the worm," was one of the best aids in reconstructing each new edition of the code as it went into effect.

SECTION XXII.

SOLUTION BY COMPARISON OR ANALOGY.

	Paragraph.
Nature of the method.....	102
Example of solution as a result of an inexcusable blunder.....	103

102. Nature of the method.—The method of solution or reconstruction by comparison or analogy is much more simple than the first method, solution by first principles, but it can be applied only in special cases, and usually only after solution by the first method has given a good insight into the nature of the messages passing between the correspondents. Solution by this method simply involves the study of a series of unsolved messages, the usual form and general contents of which are known from previously decoded messages, or, very rarely, from captured plain-text messages for which the equivalent code text is available. All that is necessary is to find the code groups in the unsolved messages in the new code which are

analogous in position and in probable meaning to the corresponding groups in the solved messages in order to form a basis for the assumption of values for the new code groups.

The greatest source of such material is to be found in messages which take the form of routine reports, such as morning and evening reports, or daily schedules of activities, because such messages very quickly assume a stereotyped or set form and phraseology. Furthermore, such messages are usually sent at definite times during the day, a factor of considerable assistance to the enemy in finding and isolating such messages from the other messages in the day's traffic. Once they have been found, comparison with solved messages of similar nature in a previous edition of the code soon leads to solution of the new code. It is obvious, therefore, that the transmission of such messages by any agency which is susceptible of interception should be strictly prohibited.

Solution by comparison or analogy may, however, find its data from sources other than that of stereotyped reports. Two actual instances will be cited. It was a regular procedure on the part of the German Army Signal Officer to call in the old editions of the code book within a few hours after the new edition went into effect. Directions were given *by radio* and *in the new edition of the code* that the old code books must be returned to such and such a station. Now it is a curious and noteworthy fact that the German code book did not contain the phrase "code book," so that any reference to it had to be spelled out. The Germans called their code book a "satzbuch," literally, "sentence-book," and when this word had to be spelled out it required from six to nine or ten groups, depending upon whether the singular or plural was used and upon the particular combinations of syllables or letters used. Therefore, when a new code went into effect it was usual to search among the messages for the first few hours' activity for a message which gave indications of being one of these particular messages, and there would usually be three or more of them, for the same message had to be sent to several stations within jurisdiction of the army and division. Once a message containing this word was found, its solution was speedily attained, with the result that the values for at least six important spelling groups were attained at one blow. How the Germans overlooked this detail seems hard to understand. Had there been one code clerk who was conscientious and intelligent in his duty he would have reported the omission of this important word from the vocabulary for his own benefit, for it would have saved him considerable time and labor if he had been able to use one code group to represent the word "satzbuch" as against being under the necessity of spelling out the word and using six or more.

103. Example of solution as a result of an inexcusable blunder.—The other instance was of far greater importance, for it led to the breaking down of an entirely new code of a type never used before and that, as a result of the interception of only three messages. On March 10, 1918, messages of an obviously new type of code (not merely a new edition of an old type) appeared along the entire Western Front. The intercepted messages of the first day's traffic along the Verdun Front, March 11, were turned in by the radio intercept stations, and among these first messages there appeared the following:

AN v X2 (Intercepted at Souilly at 0040) 0025 CHI-13
845 422 373 792 240 245 068 652 781 245 659 659 504

On the same telegram there also appeared the following:

X2 v AN (Intercepted at Souilly at 0052) 0025 CHI-13
OS RGV KZD

The second message was recognized as containing two groups which belonged to the other code, which already had been solved, though not in its entirety. The two letters OS were recognized as being a service abbreviation for the phrase "Ohne Sinn," i. e., "Your message is unintelligible." The group RGV meant "old" and the meaning of KZD was not known. After a few minutes thought this train of reasoning was followed: Station X2 sends a message to Station AN in the new code. Station AN replies twelve minutes later—"Your message of 0025 o'clock, containing 13 code groups (is) unintelligible. (Send in) old KZD." Is it not possible that KZD means "code," or "edition," or "encipherment"? Would station X2 be so foolish as to accede to such a request? A search was made immediately among the same day's messages in the letter code for a message from Station X2 to Station AN at about 0025 o'clock, and the following message was found:

AN v X2 (Intercepted at Souilly at 0057) 0025 CHI-14.
UYC REM KUL RHI KWZ RLF RNQ KR D RVJ UOB KUU UQX UFQ RQK

This message was at once decoded and found to read as follows:

UYC REM KUL RHI KWZ RLF RNQ KR D RVJ UOB KUU UQX UFQ RQK
An (?) Bn. 2 (?) h i r sch (?) w i t t e

It seemed almost too much to expect that the message in the new code should be in almost exactly the same form as this decoded message in the other already reconstructed code. Still, such was the case. Note below the internal evidences in the messages as to their similarity; the repetition of the group 659, which stands for "t"; the equivalent in the other code is the group UFQ, which stands for "tt."

Note the repetition of the group 245, which stands for "i," whereas in the letter code two different groups were used for this same letter. We may place these two messages beneath each other for comparison.

```

UYC REM KUL RHI KWZ RLF RNQ KRQ RVJ UOB KUU UQX UFQ RQK
An (?) Bn. 2 (?) h i r sch (?) w i t t e

845 422 373 792      240 245 068 652 781 245 659 659 504
An (?) Bn. 2      h i r sch w i t t e

```

From the few clues offered by this solution the nature of the entire code was speedily disclosed. Here we have a striking example of two very important things in code work. First, it shows how the solution of code messages may be attained by the method of comparison or analogy, and, secondly, it very forcibly illustrates the great danger of the violation of one of the most important rules in all cryptographic work, which rules are summarized in paragraph 106, below.

SECTION XXIII.

MISCELLANEOUS CONSIDERATIONS.

	Paragraph.
Comparison of code and cipher systems.....	104
The importance of the time element in the analysis of military cryptograms.....	105
Fundamental rules for safeguarding cryptograms.....	106
The correction of mutilated messages.....	107

104. Comparison of code and cipher systems.—Each of these two general methods of secret communication has its place in the military service, and both are at present indispensable adjuncts to any real system of signal communication. When and if cipher machines that will meet every requirement necessary in a cryptographic system for use in governmental affairs are finally invented and constructed, it may be that cipher will entirely supersede code in military, naval, and diplomatic correspondence. But so far, no single machine has yet been constructed which will meet all the requirements of simplicity, secrecy, and portability, so that it can be used for all forms of secret communication necessary in the military service. Hence, in the comparisons which follow, only cipher methods operated without machines, or in other words, "hand methods" will be considered.

The principal factors to be taken into account in comparing code and cipher methods as systems of secret communication are—

1. Simplicity, rapidity, practicability.
2. Secrecy.
3. Accuracy.
4. Economy.

1. In general it may be said that code is a more rapid, simple and practicable method of secret communication than is cipher, both as regards encoding and decoding. The processes of enciphering and deciphering require very close mental attention to avoid errors, and are usually much slower than those of encoding and decoding, which more nearly approach automatic processes and thus require less concentrated mental effort. This is of greatest importance in the combat zone, where time is most pressing, and the mental strain and excitement of battle are apt to lead to many errors. What has been said here applies, of course, only to cipher methods operated by hand, and not to ciphers produced by an automatic machine or device. There are, it is true, very small cipher devices which tend to reduce the mental strain to a minimum, but in general, the cryptograms they yield are not secure, especially when many messages are available for interception by the enemy.

2. Code systems are, on the whole, more secret than cipher systems, depending upon (a) the extent of the vocabulary and its arrangement; (b) the extent to which the code is used, that is, the number of messages transmitted. Furthermore, the solution of one message does not entail the immediate breakdown of the whole system, with the consequent solution of all other messages in the same key, as is the case in ciphers. On the other hand, in the case of code it is absolutely necessary to guard at all times the code book, so that it does not fall into the possession of the enemy. Actual possession is not always necessary, for unauthorized sight of one code, with opportunity to copy or memorize certain portions of it, is sufficient to compromise the whole code. Small codes may be carried about very easily, but then they are all the more likely to fall into the hands of the enemy. In the case of large and bulky codes, which can not be carried about so easily, very often messages can not be decoded during important movements because the books are locked up in safes or boxes which are on the road somewhere, in course of transporting them to new headquarters.

3. On the whole, it may be said that code systems are less accurate than cipher systems and are more subject to the necessity for repetition of messages, than are cipher systems. This is because a mistake in one or two code groups may obscure, alter, or render unintelligible the meaning of a whole message, whereas in the case of ciphers, the meaning of a few letters which are in error may be supplied by the context.

4. Since code text is usually shorter than the equivalent plain text, on account of the abbreviating features of code, the latter is more economical than cipher. This is of great importance where the amount of traffic is very heavy, and each unnecessary character transmitted occupies the time of a large personnel and a great amount

of equipment. On the other hand, it is true, of course, that codes must be prepared, printed and distributed, processes which take much time and labor, and are often attended with considerable difficulties. A continuously operative code compilation section must be maintained to replace codes as fast as they become compromised by continued use, or by capture. The handling of the manuscript, proofs, etc., in printing entails the necessity of ever watchful secrecy; and finally, the difficulties of a prompt and thorough distribution of the codes to all who must use them are sometimes very great, especially where this distribution must be made over an extensive territory. In the long run, therefore, ciphers are possibly more economical than codes, but this increase in degree of economy is probably very low.

105. The importance of the time element in the analysis of military cryptograms.—The question of the time necessary for the enemy cryptanalysts to reduce a cryptogram is of very great importance in a proper discussion of military cryptography. The time required for solution and the effect the solution will have on the situation can be summarized as follows:

1. Time necessary to transmit the intercepted message to the solving headquarters.
2. Time necessary to solve, including that required in making copies, tables, records, etc.
3. Time necessary to transmit the information obtained to the headquarters concerned in directing operations, including that required by the intelligence section to decide upon the authenticity of the message, its consistency with information obtained from other sources, etc.
4. Time necessary to transmit the orders determined by the information thus obtained to the combat units concerned.

Of these four elements, the only one which is subject to the greatest variation is the second, and this will depend upon many factors. First, of course, is the nature of the cryptographic system involved; secondly, the amount of text available; thirdly, the number and skill of the experts and assistants employed; and fourthly, the special conditions existing in the messages to be solved.

The purpose of this book is to show why certain systems are rapidly reduced, and the methods employed in the analysis. It will give the student a better insight into the reasons for rejecting certain methods and adopting the ones now in use. As to the second factor, the amount of material available for study, this depends entirely upon the tactical situation. Now if a system that resists analysis unless many messages are available is being used, then it

may often happen that days will go by without any solutions because of the lack of sufficient material for study. But when the activity increases, then more messages are available and the chances for solution are better. This is all the more important in that when the activity is low, the situation is more or less unimportant, but as soon as the activity increases, action may be expected, and the value of the solution of the messages will be all the greater. The fact that several days of painstaking labor may prove to be unsuccessful is counterbalanced by the fact that a single solved message may be of greater importance than all the messages of the preceding days that might have been solved but were not. The failure to solve the messages of one day or of several days does not indicate by any means that the system is indecipherable. Of course, the height of perfection as regards the security afforded by a cryptographic system would be that a single message or a great many messages all in the same key or in the same code should remain absolutely insoluble forever, but that lies outside the realm of possibility so far as our present methods in military cryptography are concerned. The best that can be expected is that the system should be complicated enough to resist analysis for such a length of time that when solution is finally achieved, the information obtained is of no special value.

With regard to the skill of the experts employed, this, of course, requires experience and training. Cryptanalysts can not be developed in a few days; moreover, not all individuals can become expert, for the peculiar nature of the science requires a person with a correspondingly peculiarly developed mind, whose principal characteristic is the ability to reason inductively and deductively, to devise methods of attack, and to persevere until success attends his efforts. Imagination, a comprehensive vocabulary, and a good knowledge of the language of the enemy, together with an indefinable element best described by the word *flair*, and finally, *luck*, all play their part in the making of an expert.

As to the special conditions existing in the messages being studied, this refers to the specific contents of the text, the number and nature of the repeated words, proper names, etc., in short, all the more or less extraneous circumstances which surround the cryptograms, and may offer clues to the alert cryptanalyst.

106. Fundamental rules for safeguarding cryptograms.— There are a few fundamental rules which must be observed in all cryptographic work. Failure to observe such rules will inevitably lead to a more rapid solution by the enemy than would otherwise be the

case. Much of the success which attends the efforts of the cryptanalyst is due to ignorance and carelessness on the part of the clerks who are entrusted with the work of encoding or enciphering messages. The following general rules would seem to be self-evident, but they are violated every day.

(a) A message once transmitted in one form or type of code or cipher must never be repeated in any other form, key, or type of code or cipher whatsoever. If, for some unknown reason, a message which has been verified and repeated is still unintelligible, or, what is more often the case, can not be decoded or deciphered, then it is necessary to *paraphrase* the message—that is, rewrite it so as to change its original wording as much as possible without changing the meaning of the message. This is done by altering the positions of sentences in the message, by altering the positions of subject, predicate, and modifying phrases or clauses in the sentence; by altering the diction without loss of sense; by deletion rather than expansion of the wording of the message. If an ordinary message is paraphrased simply by expanding it along its original lines, an expert can easily reduce the so paraphrased message to its lowest terms, and the resultant wording will be practically the original message. For this reason, deletion, if possible, is better than expansion. After paraphrasing, the message can be sent in the other key or code. So far as possible, no information of any kind should ever be given in a plain-text, code, or cipher message which may connect it in any way directly by verbiage with a message previously sent.

(b) A message once sent in code or in cipher must never be repeated in clear under any circumstances. Vice versa, a message once sent in clear must never be repeated in code or in cipher, and, of course, a code or cipher message must never be answered in clear.

(c) Never insert or leave unenciphered or unencoded plain text of any sort in code or cipher messages. This includes punctuation and abbreviations of any description. They afford valuable clues to the enemy.

(d) Plain text and its equivalent code or cipher text must never appear on the same sheet of paper for final copy or for filing purposes. Work sheets should be destroyed by burning.

(e) All rules and precautions set forth in the instructions to the various codes and ciphers issued for use must be observed very carefully. These rules have been adopted as a result of experience gained in solving enemy messages during the late war and are intended to delay the solution of our own messages as long as possible. Practice in the preparation of code messages is especially recommended be-

cause of the familiarity that is soon gained with the particular words and phrases contained within the book. With familiarity of contents, and speed in operation, the length of the messages may be reduced very considerably, as well as the time necessary to prepare them.

(f) The more messages sent, and the longer the messages are, the sooner will the enemy be able to solve them. Messages can be materially shortened by the deletion of unnecessary words, punctuation, etc. The formation and adoption of fixed habits as regards the phraseology of messages, arrangement of their contents, use of punctuation, etc., is a most dangerous practice, and will assist the enemy cryptanalysts very greatly. Routine reports of all kinds should be sent by means and agencies not susceptible of interception.

(g) Finally, the utmost care should be taken to prevent the loss or unauthorized sight of the codes or lists of cipher keys in use. It is possible to photograph an entire code in two or three hours. Mere continued possession of the code is, therefore, no absolute guaranty that it has not been compromised by photography or some other method of reproduction. The only absolute assurance of its not having been compromised is that it has never left the possession of the person into whose care it has been entrusted or the safe in which it is kept when not in use. Even if knowledge that a code has been compromised follows immediately after such compromise, the time and difficulties attendant upon the notification of the fact to all concerned and the distribution of a new code are so great that much serious damage is caused by the delay and interruption in communication, not to speak of the danger resulting from the decoding of the most recent messages in the compromised code.

107. The correction of mutilated messages.—Errors in the execution of all the operations involved in cryptographic communication are so common and are so troublesome that many commanders, who, for the most part, are already prone to regard the operations of cryptography as being hopelessly slow and cumbersome, are very much prejudiced against the employment of either cipher or code in the field of operations.

Not much can be done in the way of facilitating and making more rapid the work of enciphering or encoding and deciphering or decoding; they are admittedly slow processes, except in the case of certain cipher machines, which usually can not be carried about easily and are thus not available for use in the field. The only thing that can be done is to employ those systems which combine to the greatest extent the elements of safety and facility.

However, as regards the errors which are inevitable in such work, training and experience in the operations will greatly reduce the time necessary to correct such errors as are most commonly encountered. Such errors may be traced to four sources:

1. In enciphering or encoding, including copying.
2. In transmission by any agency other than courier.
3. In reception.
4. In deciphering or decoding, including copying.

Of these, the first and last can be eliminated by reducing the steps in cryptographic operations to a definite system and *invariably checking the work*. Great care in the formation of the letters in writing must be exercised, and Roman capitals should always be used. Whenever possible the cryptogram should be deciphered or decoded by a person other than the original encipherer or encoder as a test of its correctness before turning it over to the transmitting agency, and at the receiving end, it should be carefully checked against the original work sheets before being turned over to the addressee and before destroying these work sheets.

The second and third sources of error are harder to avoid, especially in transmission by radio telegraph, on account of interference, atmospheric disturbances, and the like. The code or cipher clerk should familiarize himself with the telegraph alphabets and the most common errors encountered in telegraphic transmission, so as to be able to refer an error to its probable cause or to find clues for the correction of the garbled groups. Such a table is given herewith. (Table 6.)

Every message should be examined as to its correspondence with the word check or letter check, as given in the preamble to the message; furthermore, each group should be examined to see that it has its proper quota of letters—no more and no less. In ciphers, this is especially important, for the omission or addition of a single letter will often render the message unintelligible.

TABLE 6.

Continental Morse alphabet (used in radio, cables, and outside United States).			American Morse alphabet (used in the United States, except for radio).		
Letters and figures.	Alphabet.	Frequent errors.	Letters and figures.	Alphabet.	Frequent errors.
A	. —	i, m, t, et	A	. —	i, t, et
B	— . . .	d, ts	B	— . . .	d, h, ts
C	— . — .	f, k, j, r, nn	C	. . .	s, z, ie
D	— . .	b, s, l, ti	D	— . .	b, ti
E	.	t, a, i	E	.	t
F	. . — .	q, r, in	F	. — .	r, q, en
G	— . — .	m, n, o, q, me	G	— . — .	n, c, me
H	s, v, b, se	H	s, p, z, y, es
I	. .	a, n, s	I	. .	a, o, e
J	. — — —	w, o, eo, am	J	— . — .	c, k, ke
K	— . —	a, n, d, o, ta	K	— . —	j, n, ta
L	. — . .	x, r, d, ed	L	—	t, n
M	— —	a, n, i, tt	M	— —	n, a, tt
N	— .	i, m, t, te	N	— .	o, t, te
O	— — —	g, k, m, w, mt	O	. .	n, i, ee
P	. — — .	j, w, g, l, r, an	P	h, s
Q	— . — . —	g, k, o, x, z, ma	Q	. . — .	f, g, u, in
R	. — . .	a, n, f, g, s, l, w	R	. . .	s, i, ci
S	. . .	h, d, i, r, u, v	S	. . .	h, r, i
T	—	a, e, n	T	—	l, e, n
U	. . —	a, s, v, it	U	. . —	v, a, w, it
V	. . . —	h, u, x, st	V	. . . —	u, st
W	. — —	a, m, o, r, u, at	W	. — —	f, a, u, m, at
X	— . . . —	d, v, u, k, y, tu	X	. — . . .	l, y, f, ai
Y	— . — — —	x, w, k, c, nm	Y	h, ii
Z	—	b, d, g, q, mi	Z	h, c, se
1	. — — — —	0, 2	1	. — — .	p
2	. . — — —	1, 3	2	. . — . .	3
3	. . . — —	2, 4	3	4
4 —	3, 5	4 —	3
5	4, 6	5	— — — —	
6	—	5, 7	6	p
7	—	6, 8	7	—	
8	—	7, 9	8	—	
9	—	8, 10	9	—	x
0	—	9	0	—	L

Errors frequent in both systems:

- . is omitted.
- is omitted.
- . is transmitted as —
- is transmitted as .

SECTION XXIV.

BIBLIOGRAPHY.

Paragraph.

Bibliography..... 108

108. Bibliography.—The number of books and papers devoted to cryptography is very limited, and most of those which are available are in foreign languages. However, in only a few of the published works will the student find explanations of the more complex types of ciphers and methods for their solution. The 15th, 16th, and 17th Centuries were most prolific in the production of works on cryptography, and many of them are very interesting. But they are only of historical importance and are very rare. The following list gives the titles of only those publications which may be considered to have a bearing upon modern cryptography:

- BALL, W. N. Rouse, *Mathematical Recreations*, pp. 395–423. London, 1917.
- BAZERIES, Etienne, *Étude sur la cryptographie militaire*, Paris, 1900.
Chiffres de Napoleon, Fontainebleau, 1896.
Les chiffres secrets dévoilés, Paris, 1901.
Le Masque de Fer, Paris, 1893.
- BLAIR, William, *Cipher*, Ree's Cyclopeda.
- DELASTELLE, F., *Cryptographie nouvelle*, Paris, 1893.
Traité élémentaire de cryptographie, 1902.
- DELAGE, Emile, *La chiffrage-cryptographie, L'Art de S'Écrire en Secret Absolu*, Paris, 1900.
- FLEISSNER, Von Wostrovitz, *Handbuch der Kryptographie*, Vienna, 1881.
- FRIEDMAN, W. F., See under Riverbank Publications.
- GIOFFI, L., *La Crittografia*, 1897.
- GRANDPRÉ, A., *De Cryptographie pratique*, Paris, 1905.
- GROSS, H., *Handbuch fuer Untersuchungsrichter*, Teil II, Munich, 1914.
- HITT, Lieut. Col. Parker, *Manual for the Solution of Military Ciphers*. Leavenworth, 1916 and 1918.
- JACOB (Le bibliophile), *La Cryptographie*, Paris, 1858.
- JOSSE, H., *La Cryptographie et ses application à l'art militaire*, Paris, 1885.
- KASISKI, F. W., *Die Geheimschriften und die Dechiffirkunst*, Berlin, 1863.
- KERCKHOFFS, A., *La Cryptographie militaire*, Paris, 1883.
- KLUEBER, J. L., *Kryptographik*, Tuebingen, 1809.
- LANGIE, André, *Cryptography*, New York, 1922.
- LOHR, Capt. Lenox R., see under Riverbank Publication No. 19.
- MAUBORGNE, Maj. J. O., *An Advanced Problem in Cryptography and its Solution*. Leavenworth, 1914.
- MEISTER, Aloys, *Die Anfaenge der modernen diplomatischen Geheimschrift*, 1902.
Die Geheimschrift in Dienste der paepstlichen Kurie, Paderborn, 1906.
- MOORMAN, F. N., *Code and Cipher in France*, Infantry Journal, p. 1039, 1920.
- RIVERBANK PUBLICATIONS. Papers (except No. 19) by W. F. Friedman, Department of Ciphers, Riverbank Laboratories, Geneva, Illinois:
A Method of Reconstructing the Primary Alphabet. No. 15, 1917.
Methods for the Solution of Running-Key Ciphers. No. 16, 1918.
An Introduction to Methods for the Solution of Ciphers. No. 17, 1918.

- RIVERBANK PUBLICATIONS. *Synoptic Tables for the Solution of Ciphers, and a Bibliography of Cipher Literature.* No. 18, 1918.
Formulae for the Solution of Geometrical Transposition Ciphers. No. 19, 1918. By Capt. Lenox R. Lohr, with an introduction by W. F. Friedman.
Several Machine Ciphers and Methods for their Solution. No. 20, 1918.
Methods of Reconstructing Primary Alphabets. No. 21, 1919.
The Index of Coincidence and Its Applications in Cryptographic Analysis. No. 22, 1922.
- ROMANINI, C. F. Vesin de, *La Cryptographie dévoilée*, Paris, 1857.
SCHNEIDER, L., *Cryptographie à l'usage de Armée*, Paris, 1912.
SCHNEICKERT, HANS, *Die Geheimschriften im Dienste des Geschaefts-und Verkehrslebens*, 1905.
VIARIS, HENRI, *L'Art de chiffrer et déchiffrer les dépêches secrètes*, Paris, 1893-1895.
VALÉRIO, P., *De la Cryptographie*, Part I, Paris, 1893.
De la Cryptographie, Part II, Paris, 1896.
WHEATSTONE, CHAS, J., *Scientific Papers of Sir Charles Wheatstone*, published by the Physical Society of London, 1879.
LITERARY DIGEST, pp. 46-51, Nov. 3, 1917.
SATURDAY EVENING POST, pp. 24-25, March 10, 1917.
MUIRHEAD, F., *Lectures in Technical Conferences of the U. S. Army Signal Schools*, 1911 to 1912, 1912 to 1913.

APPENDIX.

PROBLEMS GIVEN DURING COURSE AT CAMP VAIL,
DECEMBER 11 TO 23, 1922.

(Solutions will be found at the end of this appendix.)

PROBLEM 1.

I. Solve the following:

FTCZQ POFHM ATPOZ WDZUC HUUQJ TUQZE BDUTQ
 OADHP OTCBN WEDUP KATPO ZWDFH MHWQJ TUAZW
 WCZMD PZKOL THUEZ UQHMZ UDUTQ OAOAD QDTCK
 HVVTM ZUBOT QTHEY NUFOZ TUCZM DCZMD UHNBA
 OKKGH PFTVG

- II. What kind of a cipher is this? (Class and type.)
 III. Name two methods of solving ciphers of this type.
 IV. What is meant by the expression "spatial relations and linear extensions of the crests and troughs" in a frequency table?

PROBLEM 2.

I. Solve the following:

PRCYR NSMLN CDKKL GADXR TKXZI LZZRY
 APAUV JLCUP YDPGD MAELQ CNLBC UDBMJ
 QXKDZ XZKLS PRIAV KFXIL YKNAD NBANM
 KBYHW UBIAI PBJAH UFOCX BVQUV WXOIL

- II. What is the key word for the message?
 III. Why does the frequency table for this message show no marked crests and troughs?

PROBLEM 3.

I. Solve the following:

TSLOV HQGDS TGYIW KYMRX QAEQT CVFMM IABBD D

- II. How was this message enciphered?
 III. Why is it so easy to solve such a message? Suppose the alphabets were mixed alphabets produced by sliding a mixed sequence against the normal, and suppose you had the mixed sequence, how would you solve such a cipher? Suppose you did not have the mixed sequence?

PROBLEM 4.

I. Solve the following:

WGFEN	UQEKR	ZKVIT	UUXEK	UWAEB	XEVEP	NUADU
ANEKA	ZRWUK	UMAWE	ERZKG	UKGNY	GEEPN	UADUU
KUCJM	KRGEA	UXEAG	XAEWA	UNNEV	MRDMK	QUFYU
KZGNG	MAGNI	YEORA	UCMZK	RUAVE	AXAUN	UKGIY
MDUMO	OEPNU	ADUAN	WZDUG	YUZAM	GGUKG	ZEKGE
OEQMG	ZKWYE	NGZOU	PMGGU	AZUNM	KRZKV	MKGAJ
XENZG	ZEKNP	UVEAU	MGGMQ	TNGMA	GNITU	UXPMG
GMOZE	KQECC	MKRUA	NVBOO	JZKVE	ACURM	KRCMT
UMAAM	KWUCU	KGNVE	ANUKR	ZKWZK	VEACM	GZEKX
AZNEK	UANMK	RREQB	CUKGN	GEAUW	ZCUKG	MOQEC
CMKRX	ENGIM	QTKEF	OURWU	CUKGA	UHBUN	GURIF
ZONEK	QEO					

II. What type of cipher alphabet is used in this cryptogram? What is the key word upon which it is based? What characteristics should such a key word possess?

III. Explain the principle upon which the classification and separation of the high-frequency consonants and vowels is made possible from a study of the frequency table.

IV. Why does the frequency table exhibit crests and troughs in this cipher? Are the spatial relations of the crests and troughs in this frequency table different from those for a message enciphered by a direct or reversed normal alphabet? What determines their spatial relations in this frequency table?

V. How can the mixed alphabet you have just reconstructed be used to solve subsequent messages written by means of it when it is used as a sliding alphabet?

PROBLEM 5.

I. The following reply to the message given in Problem 4 was also intercepted. Decipher it.

YGISY GNZKN TQVYX NGPHS ZOSRG LFNLA WFSIK QAVFJ KNQOQ
JPG

PROBLEM 6.

I. The following message also uses the same basic mixed sequence as do Problems 4 and 5, but according to a slightly different system. Solve it and determine the key word applying to the message.

ZHXWL SCDFI AZRXE FYHAJ RCABN QQMEH DFPNS IHDMK RTKVQ
HAMQY QEVEI MMBZP WVRNF IBYWL ZLHXW SCDFI BLC

PROBLEM 7.

I. Solve the following:

IWEKA	DWRUB	KLTXW	ZNMGL	WJMNf	BNTTP	XAADX
SJNDT	PRKNV	IWAGA	WOAJU	KXFfP	QMCBU	AVXBM
YHEGL	OAGPS	TOEAX	VTJHT	QTNDM	OEBHG	OOAAC
JTABl	AQHqV	QSPXL	UXBKR	ZYOfP	AWWXQ	LJHNW
QOAGI	SNTCJ	TABVH	NYTOV	JCATJ	GOZxN	IWLCH
MAKLQ	DSMOB	UAVXB	MYHEG	IVZBX	DSUYA	HWUAG
IKQEC	JTABF	BZHKO	NNNBA	WWXPD	RCHTX	GZBUS
MQRWG	QKAVW	ENVVJ	QABWK	BBQMV	JAADC	KKUDC
JZTPV	WOTFM	RIXBB	ATDYB	ARTTY	AMFBT	ZGDQU
FXLJB	GKTGR	ZVUBJ	VUKNY	XLHLS	GKXHA	AYUOU

UW

II. Explain the basic principle of "factoring." Do the intervals between the repetitions in a single mixed alphabet cipher show factors?

III. How can you distinguish between a single alphabet cipher and a multiple alphabet cipher?

IV. Why is a cipher of the type you have just solved easy to analyze?

V. Is the compilation of frequency tables absolutely necessary for the solution of this type of cipher? Can you apply the method of completing the alphabet sequence?

PROBLEM 8.

I. The enemy is known to be using the system you have just seen exemplified in Problem 7, I. Information reaches you to the effect that a hostile force has taken Hunterstown and has established its headquarters in an office building there. The following message was the first one sent from the new command post to that of its superior unit. Solve it by guessing a word in the cryptogram as suggested by these circumstances.

PUYAX BLIWL UYAPN YPVLL RGHCZ HPGGS LBKQL EWPON
POQAQ YHZRS APZHB LEH HU VMWUL IUZEH PQBUD

II. You have noted that the use of a repeating key renders the analysis of a cryptogram a fairly simple process. The following cryptogram is an example of encipherment by the use of a continuous or nonrepeating key, the text of a book forming the successive key letters for encipherment. The particular book used is unknown to you but you feel sure it is in English. The message is suspected to contain the word CAVALRY. Solve it.

MTMAA RDLIH AALQG TZMUA NIOVB FCCMO OEW

148

ELEMENTS OF CRYPTANALYSIS.

PROBLEM 9.

I. Solve the following:

NGBZG IPUCG MYANX YRTHR PSYTA KLPOH GZVJJ SZAGN IXDQT
 VIQAN SHJES GJMZX QPRLZ RHTVR HEWZA XCBLM AMERO YDMLQ
 XNGBI EARLE EZQXP HHKSE WZQEH RQQZU WBHJU JUABD NRVSQ
 FARQW DPKRK TPDQZ ARQXF KXMAU YJBTM GWZQE HCJAZ SQIGT
 VVUUF JTVRO KJGQI JUTBU AYGUQ GETVR FMTVY XUJZH JETRB
 KDRGP GLKXA KLPOC RURIP IMLVY YDJYR THQVH HTSTQ JJTAX
 IQJSQ BFKAK LHBVN ULDQZ MIVQO BLSEY RLXLA JELAB AGYQP
 JXTSQ QCCEMT AKLPO UYDEE FQHXJ TVRLE ZMJCG XMRZE ENNKL
 VTVRB KHFPPI PIPTA TQVOS UARQJ SJBIE ARLEE VAYBL KPARQ
 XFKXM AUYJB TMGWQ MYRTH RPSYT AKLPF XBFQE GCKZM IYDJT
 ERISQ LBAGO IMFEL IIMQB BTMIY DJATO EURBD HJSIP IYTSG
 URNBB USIXO HFVEP ZUKTC RARNE YDVAG NBAGO LYJSU ZUEQN
 KXJIX WOHFV AUXNG BZAAR XZLII IJOBA GOLYJ QTTEW QMUYM
 TWNBR LVNXU SRIRG WZQKK RQYGH UERQF ROZTI EMQIK GTAOW
 USDDK ANNKX XNQXD QZRGU QNGBU KYRZZ LVOLP BKKAK LPYHH
 JEFSE MRWHQ PNSUA RQND D BVOXR OUGGN XPOHF VAUWU PWLIX
 WDBLX I

PROBLEM 10.

I. Solve the following:

52183 80928 91708 73452 83150 71833 94591 78663 21510
 21296 22526 71509 38871 39510 35756 12650 48993 02859
 08267 24985 41360 36795 11193 52291 40133 51764 21351
 61277 95520 99900 76960 19522 72244 01201 26515 22396
 66133 21564 30426 37167 95875 12625 26713 07107 69243
 96742 26501 35467 00911 81747 23760 56114 63472 11581
 52669 15105 10238 72171 72189 50221 63283 01626 71523
 97635 19962 29140 30729 50062 09396 78917 47236 30891
 71182 15087 38016 61307 12279 61532 96435 14993 05211
 42296 71056 267

II. How would you solve a cipher of this type in which the numbers did not follow in sequence with the letters of the alphabet, but were all mixed?

PROBLEM 11.

I. Solve the following:

First second machine gun brigade slope attack hill to battalion
 advancing position five against hill five on with western two five
 four supports nine.

II. Solve the following:

EGDSH VETAO NSOTS RRIVS DUHEN QITAA ESRWE
 EISCA RORTE OAISS IDNES NDFTF AERIS EHUWR
 TRDLO OENIR ILTSA IEPDY GAOTR EDSOS FOVTR
 IRETE DGLAA XAEDI XLNOE NEMOC RNESE SHCCO
 HNFNN TTLEE OEDSI TTBPI EEUSH TRRTE ONAVT
 KRARI IECRI JELRE OEONI NHGEF ENCPT AAAEI
 WRNOL IELOE RNMRE IFSDH CNIOE CMRIS DGHWL
 AAFVN FRAIE DZLWR NSCYC HNHTS OI

III. How do you distinguish a transposition from a substitution cipher?

PROBLEM 12.

I. The following messages, all of exactly the same length, were intercepted the same day, and are presumably in the same key. Solve them.

1. IAALN EOFSG TOGVE RANOL NDUOD EIHS ATFTD NRLVO
RODSW EEROR Q
2. TDNMR GREON ARIEU ETNYI TCOFE AIEUT TARDT EDNSO
EIPEC MFEAR N
3. ANELN EXEHG ILACE MEENL FXTEE EISIG AORWL LDLVV
ORDEL OCHOT H
4. EENET SLNNF TCOD OSEAI LFIGD WIAAR NOIHN LLNRF
VWLRE MRAIE A
- . RAMET MIONO DIUMA LLINK OATGT NNAIB TNHIT NIASD
RMSEC UIOVS A

PROBLEM 13.

I. Using the cylindrical cipher device and constructing your own key from a key word or a key phrase of your own choosing, encipher a message of about fifteen words, keeping a record of all your work on a single sheet of paper. Write your name upon it.

II. Write only the key word or key phrase and the cipher message on another sheet of paper, and nothing else.

III. Turn in both sheets.

PROBLEM 14.

I. Using a United States Army field message blank prepare the following for delivery to the message center:

No. 6.
MORRIDGE CG 2d Div.

1st Cav.,
HUNTERSTOWN,
2 June 22, 7:50 a. m.

Have driven enemy cavalry southeast of HAVERHILL Point. Encountered enemy column all arms marching southwest along HUNTERSTOWN-594-MCKNIGHTSTOWN road. Leading element passed 594 at 7:45 AM. YORK TURNPIKE clear. Continue reconnaissance.

MORRISON, Col.

II. Using a United States Army radio message (transmission) blank, prepare the preceding message for delivery to the radio station, using DFC 4.

PROBLEM 15.

Decode the following:

No. 6	8:30AM.	DFC4	HOBV	NAFP	YZBY
HIJA	HYJM	WAXJ	UTIR	NEOS	KURE
IWBU	SYZL	PXCJ	CYDO	ZIZR	RASJ
AFAF	YCUD	TYER	APQA	KINI	AXYZ
TEWA	VOND	JISD	OKLO	HYOH	JYXT
WIUD	MOUQ	LYE-	---	-FO	MACN
TULD	WUHN	YLIH	ZEBA	UJAD	MAOS
UGHU	DOYJ	EBUF	FIQE	DIRP	MUKB
QAOW	ZIKL	GUBB	ENTO	CYDO	OCIB
VUJQ					

PROBLEM 16.

I. You are the enemy code breakers working upon DFC 12. Your intercept stations have intercepted a considerable number of messages, and you have already made an exhaustive study of them so that you have succeeded in breaking up the text into the various classes of groups, such as numbers, spelling groups, etc. You have brought together, as listed below, all the messages which seem to have chains of spelling groups such as would be found in building up the names of places. You are to solve these groups.

The combat zone is the region included in the GETTYSBURG-ANTIETAM map, and you have before you the following list of names likely to be spelled out in these messages:

Andover Hollow.	Hilltown.
Antietam.	Hunterstown.
Arendtsville.	Logan.
Belmont Sh.	McElheny.
Benders Church.	McKnightstown.
Biglersville.	Mount Holly Springs.
Boyd Sh.	Mummasburg.
Caledonia Park.	Newchester.
Carlisle Junction.	Nicholson.
Cashtown.	Oldeston Cliff.
Center Mills.	Ortanna.
Dillsburg.	Pinegrove.
Fidler.	Seven Stars.
Gettysburg.	Spillman Farm.
Goldenville.	Stroudt Farm.
Goodintent Sh.	Swift Run.
Granite Hill.	Table Rock.
Hampton.	Woodside Sh.
Hanover.	Varney.
Harrisburg.	Venable Barns.
Heidlersburg.	York Springs.
Henderson Meeting House.	York Turnpike.
Herman.	

II. The spelling-group portions of the messages are as follows:

1. HILI BOJK UPQU VIBO DICZ ZEDE AHOL HILI
SARG OLAH
2. UVXY IDAB ADGE COSS AJAJ HIWQ
3. JYJI SEMS BEHO HILI IDAB
4. XITW QUZM OLAH JYJI OWOW DOGI
5. SOBK NYSC IVAS EROT DOKY AHOL ADGE
6. HILI UPQU ABDE
7. ADGE CUCE DYIC UDIB DOKY RACY
8. VOYB RUNZ UQPI SIOW VOCY SOBK
9. SIOW VANQ UVXY YZBY MITY COSS AJAJ HIWQ
10. SOBK NYSC IVAS KUUM IXCU XITW VANQ POPA
11. JIBS YGYG YVIR UJAD BUDO YZBY
12. FYLL LOAK UJJO
13. UDIB BIFI FYFI RUNZ IVAS
14. COSS UQPI RUBM AVZI MUON
15. CIGI UVXY YZBY DOKY
16. ADGE BIID CEAD NYNI TISJ ADGE
17. FYLL VOCY CAIH
18. ZIRS OWOW UVXY WYJQ MUON
19. RUBM EXBI AVZI HILI OWOW YMJE SOBK EROT
DOKY AHOL ADGE
20. COSS BOJK UPQU ADGE OHEF AJAJ OHEF
21. FYLL DOKY BORS ADGE COSS AJAJ HIWQ
22. NYNI FYFI EXBI UVXY UDIB OJUK AHIK
23. KYFZ HIWQ SEMS
24. UJAD TISJ IRUT SOBK
25. DYEB ARUW ADGE COSS AJAJ HIWQ
26. KOMI HIWQ YZBY MITY UJAD IDAB OLAH
27. HILI RACY YSNA NYNI UHAC
28. HILI IDAB FYZZ
29. NYSC XEMN VOCY KIEL
30. RUNZ CEAD YSNA VOYB IDAB ADGE
31. ISER MUON FYZZ
32. HIWQ VIKP CEAD UJAD BUDO YZBY
33. ISER YZBY XEXJ ATBU YQIM ZUBI DOKY IVAS
34. VOYB RUNZ IVAS KOVG ADGE FYZZ
35. ISER DOKY ATAT ADGE YZBY SAMR RACY RUNZ
LIVC
36. ABGO OWOW UVXY QUZM UDIB PEVO MUON
37. COSS OWOW SOBK UVXY MUON
38. IRUT CUCE OHEF AJLE YSNA
39. EROT VANQ YMJE ABDE OJUK AHIK
40. DOGI VOCY BIFI COSS YQIM IXCU ADGE
41. YQIM YGYG CAIH HILI PAVI KYFZ CUCE

42. VIKP AJLE UJJO RUNZ UGCA DYIC OJUK
 43. ATBU OHEF PAVI VIBO
 44. KOMI KUUM VOYB QUZM SARG
 45. APRE YQIM ATAT CIGI ISER LIVC ADGE
 46. OLAH LOAK ARUW
 47. WYJQ AHIK ADGE
 48. RUNZ VANQ HIWQ JIBS YVIR
 49. DOKY BIID ZEDE DICZ AVZI
 50. APRE BEHO ZIRS DOKY BORS EBCE

III. The following is the index of the groups occurring in these messages:

ABDE—6, 39	KYFZ—23, 41
ABGO—36	LIVC—35, 45
ADGE—2, 5, 7, 16, 16, 19, 20, 21, 25, 30, 34, 35, 40, 45, 47	LOAK—12, 46
AHIK—22, 39, 47	MITY—9, 26
AHOL—1, 5, 19	MUON—14, 18, 31, 36, 37
AJAJ—2, 9, 20, 21, 25	NYNI—16, 22, 27
AJLE—38, 42	NYSC—5, 10, 29
APRE—45, 50	OHEF—20, 20, 38, 43
ARUW—25, 46	OJUK—22, 39, 42
ATAT—35, 45	OLAH—1, 4, 26, 46
ATBU—33, 43	OWOW—4, 18, 19, 36, 37
AVZI—14, 19, 49	PAVI—4, 43
BEHO—3, 50	PEVO—36
BIFI—13, 40	POPA—10
BIID—16, 49	QUZM—4, 36, 44
BOJK—1, 20	RACY—7, 27, 35
BORS—21, 50	RUBM—14, 19
BUDO—11, 32	RUNZ—8, 13, 30, 34, 35, 42, 48
CAIH—17, 41	SAMR—35
CEAD—16, 30, 32	SARG—1, 44
CIGI—15, 45	SEMS—3, 23
COSS—2, 9, 20, 21, 25, 37, 40	SIOW—8, 9
CUCE—7, 38, 41	SOBK—5, 8, 10, 19, 24, 37
DICZ—1, 49	TISJ—16, 24
DOGI—4, 40	UDIB—7, 13, 22, 36
DOKY—5, 7, 15, 19, 21, 33, 35, 49, 50	UGCA—42
DYEB—25	UHAC—27
DYIC—7, 42	UJAD—11, 24, 26, 32
EBCE—50	UJJO—12, 42
EROT—5, 19, 39	UPQU—1, 6, 20
	UQPI—8, 14
	UVXY—2, 9, 15, 18, 22, 36, 37
	VANQ—9, 10, 39, 48

ELEMENTS OF CRYPTANALYSIS.

153

EXBI—19, 22	VIBO—1, 43
FYFI—13, 22	VIKP—32, 42
FYLL—12, 17, 21	VOCY—8, 17, 29, 40
FYZZ—28, 31, 34	VOYB—8, 30, 34, 44
HILI—1, 1, 3, 6, 19, 27, 28, 41	WYJQ—18, 47
HIWQ—2, 9, 21, 23, 25, 26, 32, 48	XEMN—29
IDAB—2, 3, 26, 28, 30	XEXJ—33
IRUT—24, 38	XITW—4, 10
ISER—31, 33, 35, 45	YGYG—11, 41
IVAS—5, 10, 13, 33, 34	YMJE—19, 39
IXCU—10, 40	YSNA—27, 30, 38
JIBS—11, 48	YQIM—33, 40, 41, 45
JYJI—3, 4	YVIR—11, 48
KIEL—29	YZBY—9, 11, 15, 26, 32, 33, 35
KOMI—26, 44	ZEDE—1, 49
KOVG—34	ZIRS—18, 50
KUUM—10, 44	ZUBI—33

PLAIN TEXT OF MESSAGES GIVEN IN PROBLEMS.

PROBLEM 1.

Co, 1st Cav.

Hostile infantry on ridge northeast of GULDENS. Hostile cavalry on hill 562 and in ravine north thereof. Am moving to road junction 550.

BASCOMB.

PROBLEM 2.

CAPT. ARMSTRONG, G-4

Direct ammunition train to close at once on artillery combat trains. Further orders for it at MUMMASBURG.

JOHNSON, G-4.

PROBLEM 3.

Casualties severe. Reinforcements needed.

PROBLEM 4.

G-2, 2D INF.

Keep one group of observers on ridge near GOODINTENT SH. to observe enemy and to report progress of advance when it starts. Hold remainder for present. Have all observers give their attention to locating hostile batteries and infantry positions before attack starts. Keep battalion commands fully informed and make arrangements for sending information, prisoners, and documents to regimental command post. Acknowledgment requested.

WILSON, Col.

154

ELEMENTS OF CRYPTANALYSIS.

PROBLEM 5.

Co, 2D INF.

Acknowledge your number six.

MARTIN, G-2.

PROBLEM 6.

G-1, FIRST ARMY.

Ten Hotchkiss machine guns and two 37 mm. guns captured.

G-1, 1ST Div.

PROBLEM 7.

G-4, 2D DIVISION.

2d Echelon leaves here at 12:15 P. M. for HUNTERSTOWN. Animal-drawn vehicles of engineer train follow field trains. Motor section, sanitary trains should arrive in the neighborhood of HUNTERSTOWN at 1:00 P. M. Ammunition train, including horsed battalion, has been sent forward in rear of artillery combat trains and should reach vicinity of GOLDENVILLE by 1:30 P. M.

WHEELER, G-4.

PROBLEM 8.

I. Co, 2D CAV.

Enemy in retreat toward NEWCHESTER. Am moving to WALLACE FARM.

REED.

II. Hostile cavalry has crossed ROCK CREEK.

PROBLEM 9.

The plan of signal communications is published as an annex to the field orders of the campaign. It prescribes the general principles that are to be followed during the campaign and the special details that continue through the several operations. The plan gives the general directions for handling the commercial telephone lines in friendly and in hostile territory, the general regulations under which the radio will operate in the nets, and matters of the like general nature. The special details of communication prescribed in the plan for the campaign contain the wave lengths and call letters assigned to each net and unit, the code of pyrotechnic signals, panel signals, the code names of units and individuals, the particular message codes, map coordinate codes and all other items in connection with employment of the agencies used in signal communication.

PROBLEM 10.

To Regimental Commanders:

Hostile reinforced brigade occupying position from HUNTERSTOWN southwest to hill 597 inclusive. Division attack early this afternoon. Regiments to positions designated in order which follows.

CLARKSON, Adj.

PROBLEM 11.

I. 1st Machine Gun Battalion advancing with 2d Brigade to position on western slope hill 552 supports attack against hill 549.

II. A hostile reinforced brigade has occupied ridge from hill 603 to hill 597 west of SQUARE CORNER and is intrenching the main spurs that project westward. Enemy artillery is located east of orchard near eastern edge of woods on hill 732 and near crossroads 591. 1st Division attacks.

PROBLEM 12.

1. Have ordered ration wagons of 1st Squadron to GOLDENVILLE.
2. Enemy defeated. Direction of retreat not certain. Am pursuing.

3. Second echelon will leave here at 8 A. M. for GOLDENVILLE.

4. Animal-drawn vehicles of engineer train follow field trains.

5. Ammunition train including horsed battalion moves at 6 A. M.

PROBLEM 15.

CHIEF OF STAFF.

Report aero reconnaissance to 7:30 A. M. Enemy unsuccessfully tri(ed) observation west of line GETTYSBURG—MOUNT HOLLY SPRINGS. HAMPTON and NEWCHESTER apparently occupied by enemy. Observation difficult.

PROBLEM 16.

G-2.

- | | |
|----------------------------|------------------------|
| 1. Henderson Meetinghouse. | 26. Biglersville. |
| 2. Dillsburg. | 27. Hunterstown. |
| 3. Granite Hill. | 28. Hilltown. |
| 4. Pine Grove. | 29. Ortanna. |
| 5. York Springs. | 30. Center Mills. |
| 6. Herman. | 31. Cashtown. |
| 7. Swift Run. | 32. Goldenville. |
| 8. McElheny. | 33. Caledonia Park. |
| 9. Heidlersburg. | 34. McKnightstown. |
| 10. York Turnpike. | 35. Carlisle Junction. |
| 11. Arendtsville. | 36. Goodintent Sh. |
| 12. Hampton. | 37. Boyd Sh. |
| 13. Table Rock. | 38. Newchester. |
| 14. Belmont Sh. | 39. Spillman Farm. |
| 15. Fidler. | 40. Venable Barns. |
| 16. Seven Stars. | 41. Andover. |
| 17. Hanover. | 42. Oldeston Cliff. |
| 18. Woodside Sh. | 43. Nicholson. |
| 19. Mount Holly Springs. | 44. Bituminous. |
| 20. Benders Church. | 45. Qualifications. |
| 21. Harrisburg. | 46. Empty. |
| 22. Stroudt Farm. | 47. Side arms. |
| 23. Logan. | 48. Cigaretts. |
| 24. Varney. | 49. Revetment. |
| 25. Gettysburg. | 50. Quite worried. |

156

ELEMENTS OF CRYPTANALYSIS.

SIGNAL CORPS PAMPHLETS.

(Corrected to Jan. 1, 1924.)

RADIO COMMUNICATION PAMPHLETS.

(Formerly designated Radio Pamphlets.)

No.

1. Elementary Principles of Radio Telegraphy and Telephony (edition of 4-28-21). (W. D. D. 1064.)
2. Antenna Systems.
3. Radio Receiving Sets (SCR-54 and SCR-54-A) and Vacuum Tube Detector Equipment (Type DT-3-A).
5. Airplane Radio Telegraph Transmitting Sets (Types SCR-65 and 65-A).
6. Loop Radio Telegraph Set (Type SCR-77-A). (W. D. D. 1115.)
9. Amplifiers and Heterodynes. (W. D. D. 1092.)
11. Radio Telegraph Transmitting Sets (SCR-74; SCR-74-A).
13. Airplane Radio Telegraph Transmitting Set (Type SCR-73).
14. Radio Telegraph Transmitting Set (Type SCR-69).
17. Sets, U. W. Radio Telegraph (Types SCR-79-A and SCR-99). (W. D. D. 1084.)
20. Airplane Radio Telephone Sets (Types SCR-68; SCR-68-A; SCR-114; SCR-116; SCR-59; SCR-59-A; SCR-75; SCR-115).
22. Ground Radio Telephone Sets (Types SCR-67; SCR-67-A). (W. D. D. 1091.)
23. U. W. Airplane Radio Telegraph Set (Type SCR-80).
24. Tank Radio Telegraph Set (Type SCR-78-A).
25. Set, Radio Telegraph, Type SCR-105. (W. D. D. 1077.)
26. Sets, U. W. Radio Telegraph, Types SCR-127 and SCR-130. (W. D. D. 1056.)
27. Sets, Radio Telephone and Telegraph, Type SCR-109-A and SCR-159. (W. D. D. 1111.)
28. Wavemeters and Decremeters. (W. D. D. 1094.)
30. The Radio Mechanic and the Airplane.
40. The Principles Underlying Radio Communication (edition of May, 1921). (W. D. D. 1069.)
41. Introduction to Line Radio Communication. (W. D. D. 1114.)

WIRE COMMUNICATION PAMPHLETS.

(Formerly designated Electrical Engineering Pamphlets.)

1. The Buzzerphone (Type EE-1).
2. Monocord Switchboards of Units Type EE-2 and Type EE-2-A and Monocord Switchboard Operator's Set Type EE-64. (W. D. D. 1081.)
3. Field telephones (Types EE-3; EE-4; EE-5).
4. Laying Cable in the Forward Area (formerly designated Training Pamphlet No. 3).
6. Trench Line Construction (formerly designated Training Pamphlet No. 6-a).
7. Signal Corps Universal Test Set, Type EE-65. (W. D. D. 1020.) (2d edition.)
11. Elements of the Automatic Telephone System. (W. D. D. 1096.)

TRAINING MANUALS.

20. Basic Signal Communication. (Students Manual.)
21. Basic Signal Communication. (Instructors Guide.)
22. Telephone Switchboard Operator. (Students Manual.)
23. Telephone Switchboard Operator. (Instructors Guide.)
24. Message Center Specialist. (Students Manual.)
25. Message Center Specialist. (Instructors Guide.)

TRAINING PAMPHLETS.

1. Elementary Electricity (edition 1-1-21). (W. D. D. 1055.)
2. Instructions for Using the Cipher Device, Type M-94. (W. D. D. 1097.) For official use only.
3. Elements of Cryptanalysis. (W. D. D. 1117.) For official use only.
4. Visual Signaling.
7. Primary Batteries (edition of 6-9-22). (W. D. D. 1112.)
8. Storage Batteries. (Formerly designated "Radio Pamphlet No. 8.")

FIELD PAMPHLETS.

1. Directions for Using the 24-cm. Signal Lamp, Type EE-7.
2. Directions for Using the 14-cm. Signal Lamp, Type EE-6.

TRAINING REGULATIONS.

(Signal Corps subjects.)

- 165-5. Wire Axis Installation and Maintenance within the Division.