

~~CONFIDENTIAL~~

314

~~CONFIDENTIAL~~

314

Declassified and approved for release by
NSA on 10-16-2014 pursuant to E.O.
13526

Record taken from
WFF's home

~~CONFIDENTIAL~~

~~RESTRICTED~~

WAR DEPARTMENT
OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

MILITARY CRYPTANALYSIS

PART III

APERIODIC SUBSTITUTION SYSTEMS

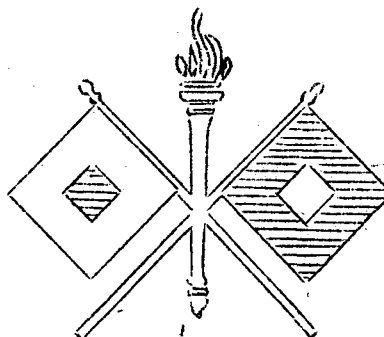
by

WILLIAM F. FRIEDMAN

Principal Cryptanalyst,

SIGNAL INTELLIGENCE SERVICE

Prepared under the direction of the Chief Signal Officer.



1938

~~CONFIDENTIAL~~

30 April 1959

This document is re-graded "CONFIDENTIAL" UP
of DOD Directive 5200.1 dated 8 July 1957,
and by authority of the Director, National
Security Agency.



Paul S. Willard
Colonel, AGC
Adjutant General

~~CONFIDENTIAL~~
MILITARY CRYPTANALYSIS, PART III

APERIODIC SUBSTITUTION SYSTEMS

CONTENTS

<u>Section</u>	<u>Paragraphs</u>	<u>Pages</u>
I. Introductory	1-4	1-8
II. Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, I	5-9	8-12
III. Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, II	10-13	13-22
IV. Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, III	14-16	22-30
V. Solution of systems using variable-length keying units to encipher constant-length plain-text groupings	17-22	31-45
VI. Solution of auto-key systems, I	23-26	46-58
VII. Solution of auto-key systems, II	27-29	59-73
VIII. Solution of auto-key systems, III	30-33	74-88
IX. Methods of lengthening or extending the key	34-36	89-93
X. General principles underlying solution of systems employing long or continuous keys.	37-40	94-101
XI. The cryptanalytic coincidence test.....	41-44	101-122
XII. The matching of frequency distributions...	45-52	123-148
XIII. Concluding remarks	53-54	149-150

SECTION I.

INTRODUCTORY

	Paragraph
Preliminary remarks	1
General remarks upon the nature of cryptographic periodicity	2
Effects of varying the length of the plain-text groupings . .	3
Primary and secondary periods; resultant periods	4

1. Preliminary remarks. - a. The text immediately preceding this devoted itself almost exclusively to polyalphabetic substitution systems of the type called repeating-key ciphers. It was seen how a regularity in the employment of a limited number of alphabets results in the manifestation of periodicity or cyclic phenomena in the cryptogram, by means of which the latter may be solved. The difficulty in solution is directly correlated with the type and number of cipher alphabets employed in specific examples.

b. Two procedures suggest themselves for consideration when the student cryptanalyst realizes the foregoing circumstances and thinks of methods to eliminate the weaknesses inherent in this cryptographic system. First, noting that the difficulties in solution increase as the length of the key increases, he may study the effects of employing much longer keys to see if one would be warranted in placing much trust in that method of increasing the security of the messages. Upon second thought, however, remembering that as a general rule the first step in the solution consists in ascertaining the number of alphabets employed, it seems to him that the most logical thing to do would be to use a procedure which will avoid periodicity altogether, will thus eliminate the cyclic phenomena that are normally manifested in cryptograms of a periodic construction, and thus prevent an enemy cryptanalyst from taking even a first step toward solution. In other words, he will investigate the possibilities of aperiodic systems

first and if the results are unsatisfactory, he will then see what he can do with systems using lengthy keys.

c. Accordingly, the first part of this text will be devoted to an examination of certain of the very simple varieties of aperiodic, polyalphabetic substitution systems; after this, methods of extending or lengthening short mnemonic keys, and systems using lengthy keys will be studied.

2. General remarks upon the nature of cryptographic periodicity. - a. When the thoughtful student considers the matter of periodicity in polyalphabetic substitution systems and tries to ascertain its real nature, he notes, with some degree of interest and surprise perhaps, that it is composed of two fundamental factors, because there are in reality two elements involved in its production. He has, of course, become quite familiar with the idea that periodicity necessitates the use of a keying element and that the latter must be employed in a cyclic manner. But he now begins to realize that there is another element involved, the significance of which he has perhaps not fully appreciated, viz, that unless the key is applied to constant-length plain-text groups no periodicity will be manifested externally by the cryptogram, despite the repetitive or cyclic use of a constant-length key. This realization is quickly followed by the idea that possibly all periodicity may be avoided or suppressed by either or both of two ways: (1) by using constant-length keying units to encipher variable-length plain-text groupings or (2) by using variable-length keying units to encipher constant-length plain-text groupings.

b. The student at once realizes also that the periodicity exhibited by repeating-key ciphers of the type studied in the preceding text is of a very simple character. There, successive letters of the repetitive key were applied to successive letters of the text. In respect to the employment of

the key, the cryptographic or keying process may be said to be constant or fixed in character. This terminology remains true even if a single keying unit serves to encipher two or more letters at a time, provided only that the groupings of plain-text letters are constant in length. For example, a single keyletter may serve to encipher two successive plain-text letters; if the key is repetitive in character and the message is sufficient in length, periodicity will still be manifested by the cryptogram and the latter can be solved by the methods indicated in the preceding text.¹ Naturally, those methods would have to be modified in accordance with the specific type of grouping involved. In this case the factoring process would disclose an apparent key length twice that of the real length. But study of the frequency distributions would soon show that the 1st and 2d distributions were similar, the 3d and 4th, the 5th and 6th, and so on, depending upon the length of the key. The logical step is therefore to combine the distributions in proper pairs and proceed as usual.

c. In all such cases of encipherment by constant-length groupings, the apparent length of the period (as found by applying the factoring process to the cryptogram) is a multiple of the real length and the multiple corresponds to the length of the groupings, that is, the number of plain-text letters enciphered by the same keyletter.

d. The point to be noted, however, is that all these cases are still periodic in character, because both the keying units and the plain-text groupings are constant in length.

3. Effects of varying the length of the plain-text groupings. - a.
But now consider the effects of making one or the other of these two elements variable in length. Suppose that the plain-text groupings are made variable

¹In this connection, see Section 3, Military Cryptanalysis, Part II.

in length and that the keying units are kept constant in length. Then, even though the key may be cyclic in character and may repeat itself many times in the course of encipherment, external periodicity is suppressed, unless the law governing the variation in plain-text groupings is itself cyclic in character, and the length of the message is at least two or more times that of the cycle applicable to this variable grouping.

b. (1) For example, suppose the correspondents agree to use reversed standard cipher alphabets with the keyword SIGNAL, to encipher a message, the latter being divided up into groups as shown below:

S	I	G	N	A	L	S	I	G	N	A	L	S	I	G
1	12	123	1234	12345	1	12	123	1234	12345	1	12	123	1234	12345
C	OM	MAN	DING	GENER	A	LF	IRS	TARM	YHASI	S	SU	EDC	RDER	SEFFE
<u>Q</u>	<u>UW</u>	<u>UGT</u>	<u>KFAH</u>	<u>UWNWJ</u>	<u>L</u>	<u>HN</u>	<u>ARQ</u>	<u>NGPU</u>	<u>PGNVF</u>	<u>I</u>	<u>TR</u>	<u>OPE</u>	<u>RFER</u>	<u>OCBBC</u>

N	A	L	S	I	G	N	A	L	S	I	G	N	A	L
1	12	123	1234	12345	1	12	123	1234	12345	1	12	123	1234	12345
C	TI	VET	WENT	YFIRS	T	AT	NOO	NDIR	ECTIN	G	TH	ATT	ELEP	HONES
L	HS	QHS	WOFZ	<u>KDARQ</u>	<u>N</u>	<u>NU</u>	<u>NMM</u>	<u>YIDU</u>	<u>OQZKF</u>	<u>C</u>	<u>NZ</u>	<u>NUU</u>	<u>WPWL</u>	<u>EXYHT</u>

S	I	G	N	A	L	S	I
1	12	123	1234	12345	1	12	123
C	OM	MAS	WITC	HBOAR	D	SC	OMM ...
<u>Q</u>	<u>UW</u>	<u>UGO</u>	<u>RFUL</u>	<u>TZMAJ</u>	<u>I</u>	<u>AQ</u>	<u>UWW ...</u>

Cryptogram.

<u>QUWUG</u>	TKFAH	UWNWJ	LHNAR	<u>QNGPU</u>	PGNVF
ITROP	ERFER	OCBBC	LHSQH	SWOFZ	<u>KDARQ</u> NNUNM
MYIDU	OQZKF	CNZNU	UWPWL	EXYHT	<u>QUWUG</u>
ORFUL	TZMAJ	IAQUW	W ...		

Figure 1.

(2) The cipher text in this example (Fig. 1) shows a tetragraphic and a pentagraphic repetition. The two occurrences of QUWUG (= COMMA) are separated by an interval of 90 letters; the two occurrences of ARQN (= IRST) by 39 letters. The former repetition (QUWUG), it will be noted, is a true periodic repetition, since the plain-text letters, their grouping, and the keyletters are identical. The interval in this case, if counted in terms of letters, is the product of the keying cycle, 6, by the grouping cycle, 15. The latter repetition, (ARQN), is not a true periodic repetition in the sense that both cycles have been completed at the same point, as is the case in the former repetition. It is true that the cipher letters ARQN, representing IRST both times, are produced by the same keyletters, I and G, but the enciphering points in the grouping cycle are different in the two cases. Repetitions of this type may be termed partially-periodic repetitions, to distinguish them from those of the completely-periodic type.

c. When the intervals between the two repetitions noted above are more carefully studied, especially from the point of view of the interacting cycles which brought them about, it will be seen that counting according to groupings and not according to single letters, the two pentagraphs QUWUG are separated by an interval of 30 groupings. Or, if one prefers to look at the matter in the light of the keying cycle, the two occurrences of QUWUG are separated by 30 key letters. Since the key is but 6 letters in length, this means that the key has gone through 5 cycles. Thus, the number 30 is the product of the number of letters in the keying cycle (6) by the number of different-length groupings in the grouping cycle (5). The interaction of these two cycles may be conceived of as partaking of the nature of two gears which are in mesh, one driven by the other. One of these gears has 6 teeth, the other 5, and the teeth are numbered. If the two gears are ad-

justed so that the "number 1 teeth" are adjacent to each other, and the gears are caused to revolve, these two teeth will not come together again until the larger gear has made 5 revolutions and the smaller one 6. During this time, a total of 30 meshings of individual teeth will have occurred. But since one revolution of the smaller gear (=the grouping cycle) represents the encipherment of 15 letters, when translated in terms of letters, the 6 complete revolutions of this gear mean the encipherment of 90 letters. This accounts for the period of 90, when stated in terms of letters.

d. The two occurrences of the other repetition, ARQN, are at an interval of 39 letters; but in terms of the number of intervening groupings, the interval is 12, which is obviously two times the length of the keying cycle. In other words, the key has in this case passed through 2 cycles.

e. In a long message enciphered according to such a scheme as the foregoing there would be many repetitions of both types discussed above (the completely-periodic and the partially-periodic) so that the cryptanalyst might encounter some difficulty in his attempts to reach a solution, especially if he had no information as to the basic system. It is to be noted in this connection that if any one of the groupings exceeds say 5, 6, or 7 letters in length, the scheme may give itself away rather easily, since it is clear that within each grouping the encipherment is strictly monoalphabetic. Therefore, in the event of groupings of more than 5 or 6 letters, the monoalphabetic equivalents of tell-tale words such as ATTACK, BATTALION, DIVISION, etc., would stand out. The system is most efficacious, therefore, with short groupings.

f. It should also be noted that there is nothing about the scheme which requires a regularity in the grouping cycle such as that embodied in the example. A lengthy grouping cycle such as the one shown below may just

as easily be employed, it being guided by a key of its own; for example, the number of dots and dashes contained in the International Morse signals for the letters composing the phrase DECLARATION OF INDEPENDENCE might be used. Thus, A (.-) has 2, B (- ...) has 4, and so on. Hence:

DECLARATION OF INDEPENDENCE
3 1 4 4 2 3 2 1 2 3 2 3 4 2 2 3 1 4 1 2 3 1 2 4 1

The grouping cycle is $3+1+4+4+2 \dots$, or 60 letters in length. Suppose the same phrase is used as an enciphering key for determining the selection of cipher alphabets. Since the phrase contains 25 letters, the complete the least common multiple of period of the system would be 25 and 60 or 300 letters. This system might appear to yield a very high degree of cryptographic security. But the student will see as he progresses that the security is not so high as he may at first glance suppose it to be.

4. Primary and secondary periods; resultant periods. - a. It has been noted that the length of the complete period in a system such as the foregoing is the least/ common multiple of the length of the two component or interacting periods. In a way, therefore, since the component periods constitute the basic element of the scheme, they may be designated as the basic or primary periods. These are also hidden or latent periods. The apparent or patent period, that is, the complete period, may be designated as the secondary or resultant period. In certain types of cipher machines, there may be more than two primary periods, which interact to produce a resultant period; also, there are cases in which the latter may interact with another primary period to produce a tertiary period; and so on. The final, or resultant, or apparent period is the one which is usually ascertained first as a result of the study of the intervals between repetitions. This may or may not be broken down into its component primary periods.

b. Although a solution may often be obtained without breaking down a

resultant period into its component primary periods, the reading of many messages pertaining to a widespread system of secret communication is much facilitated when the analysis is pushed to its lowest level, that is, to the point where the final cryptographic scheme has been reduced to its simplest terms. This may involve the discovery of a multiplicity of simple elements which interact in successive cryptographic strata.

SECTION II

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS
TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, I

	Paragraph
Introductory remarks	5
Aperiodic encipherment produced by groupings according to word lengths	6
Solution when direct standard cipher alphabets are employed	7
Solution when reversed standard cipher alphabets are employed	8
Comments on foregoing cases	9

5. Introductory remarks. - a. The system described in Par. 3 above is obviously not to be classified as aperiodic in nature, despite the injection of a variable factor which in that case was based upon irregularity in the length of one of the two elements involved in polyalphabetic substitution. The variable factor was there subject to a law which in itself was periodic in character.

b. To make such a system truly aperiodic in character, by elaborating upon the basic scheme for producing variable-length plain-text groupings, would be possible, but impractical. For example, using the same method as is given in Par. 3f for determining the lengths of the groupings, one might employ the text of a book; and if the latter is longer than the message to be enciphered, the cryptogram would certainly show no periodicity as regards the intervals between repetitions, which would be plentiful. However, as already indicated, such a scheme would not be very practical for regular

communication between a large number of correspondents, for reasons which are no doubt apparent. The book would have to be safeguarded as would a code; enciphering and deciphering would be quite slow, cumbersome, and subject to error; and, unless the same key-text were used for all messages, methods or indicators would have to be adopted to show exactly where encipherment begins in each message. A simpler method for producing constantly changing, aperiodic plain-text groupings therefore, is to be sought.

6. Aperiodic encipherment produced by groupings according to word lengths. -- a. The simplest method for producing aperiodic plain-text groupings is one which has doubtless long ago presented itself to the student, viz. encipherment according to the actual word lengths of the message to be enciphered.

b. Although the average number of letters composing the words of any alphabetical language is fairly constant, successive words comprising plain text vary a great deal in this respect, and this variation is subject to no law.¹ In telegraphic English, for example, the mean-length of words is 5.2 letters; the words may contain from one to 15 or more letters, but the successive words vary in length in an extremely irregular manner, no matter how long the text may be.

c. As a consequence, the use of word lengths for determining the number of letters to be enciphered by each keyletter of a repetitive key commends itself to the inexperienced cryptographer as soon as he comes to understand the way in which repeating-key ciphers are solved. If there is

¹It is true, of course, that the differences between two writers in respect to the lengths and characters of the words contained in their personal vocabularies are often marked and can be measured. These differences may be subject to certain laws, but the latter are not of the type in which we are interested, being psychological rather than mathematical in character. See Rickert, E., New Methods for the Study of Literature, University of Chicago Press, Chicago, 1927.

no periodicity in the cryptograms, how can the letters of the cipher text, written in 5-letter groups, be distributed into their respective monoalphabets? And if this very first step is impossible, how can the cryptograms be solved?

7. Solution when direct standard cipher alphabets are employed. -- a.

Despite the foregoing rhetorical questions, the solution of this case is really quite simple. It merely involves a modification of the method given in a previous text², wherein solution of a monoalphabetic cipher employing a direct standard alphabet is accomplished by completing the plain-component sequence. There, all the words of the entire message come out on a single generatrix of the completion diagram. In the present case, since the individual, separate words of a message are enciphered by different keyletters, these words will reappear on different generatrices of the diagram.

All the cryptanalyst has to do is to pick them out. He can do this once he has found a good starting point, by using a little imagination and following clues afforded by the context.

b. An example will make the method clear. The following message (note its brevity) has been intercepted:

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

c. Submitting the message to routine study, the first step is to use normal alphabet strips and try out the possibility of direct standard alphabets having been used. The completion diagram for the first 10 letters of the message is shown in Fig. 2.

d. Despite the fact that the text does not all reappear on the same generatrix, the solution is a very simple matter because the first three words of the message are easily found: CAN YOU GET. The keyletters may be sought

²Military Cryptanalysis, Part I, Par. 20.

- 11 -

in the usual manner and are found to be REA. One may proceed to set up the remaining letters of the message on sliding normal alphabets, or one may assume various keywords such as READ, REAL, REAM, etc., and try to continue the decipherment in that way. The former method is easier. The completed solution is as follows:

R	E	A	D	E	R	S
CAN	YOU	GET	FIRST	REGIMENT	BY	RADIO
TRE	CSY	GET	ILUVW	VIKMQIRX	SP	JSVAG

D	I	G	E	S	T
OUR	PHONE	NOW	OUT	OF	COMMISSION
RXU	XPWVM	TUC	SYX	GX	VHFFBLLBHG

e. Note the key in the foregoing case: It is composed of the successive key-letters of the phrase READERS DIGEST.

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

Figure 2.

8. Solution when reversed standard cipher alphabets are employed. It should by this time hardly be necessary to indicate that the only change in the procedure set forth in Par. 7c,d in the case of reversed standard cipher alphabets is that the letters of the cryptogram must be converted into their plain-component (direct standard) equivalents before the completion-sequence is applied to the message.

9. Comments on foregoing cases. - a. The foregoing cases are so simple in nature that the detailed treatment accorded them would seem hardly to be warranted at this stage of study. However, they are necessary and

valuable as an introduction to the more complicated cases to follow.

b. Throughout this text, whenever encipherment processes are under discussion, the pair of enciphering equations commonly referred to as characterizing the so-called Vigenère method will be understood, unless otherwise indicated. This method involves the pair of enciphering equations $\theta_{p/1} = \theta_{k/2}$, that is, the index letter, which is usually the initial letter of the plain component is set opposite the keyletter on the cipher component; the plain-text letter to be enciphered is sought on the plain component and its equivalent is the letter opposite it on the cipher component.¹

c. The solution of messages prepared according to the two preceding methods is particularly easy, for the reason that standard cipher alphabets are employed and these, of course, are derived from known components. The significance of this statement should by this time be quite obvious to the student. But what if mixed alphabets are employed, so that one or both of the components upon which the cipher alphabets are based are unknown sequences? The simple procedure of completing the plain component obviously cannot be used. Since the messages are polyalphabetic in character, and since the process of factoring cannot be applied, it would seem that the solution of messages enciphered in different alphabets and according to word lengths would be a rather difficult matter. However, it will soon be made clear that the solution is not nearly so difficult as first impression might lead the student to imagine.

¹See in this connection, Military Cryptanalysis, Part II, Section II, and Appendix 1.

SECTION III

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS
TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, II.

	Paragraph
Solution when the original word lengths are retained in the cryptogram	10
Solution when other types of alphabets are employed	11
Isomorphism and its importance in cryptanalytics	12
Illustration of the application of phenomena of isomorphism in solving a cryptogram	13

10. Solution when the original word lengths are retained in the cryptogram. - a. This case will be discussed not because it is encountered in practical military cryptography but because it affords a good introduction to the case in which the original word lengths are no longer in evidence in the cryptogram, the latter appearing in the usual 5-letter groups.

b. Reference is made at this point to the phenomenon called idiomorphism, and its value in connection with the application of the principles of solution by the "probable-word" method, as explained in a previous text¹.

When the original word lengths of a message are retained in the cryptogram, there is no difficulty in searching for and locating idiomorphs and then making comparisons between these idiomorphic sequences in the message and special word patterns set forth in lists maintained for the purpose. For example, in the following message note the underlined groups and study the letters within these groups:

¹ Military Cryptanalysis, Part I, Par. 33 a-d, incl.

MESSAGE:

XIXLP EQVIB VEFHAPFVT RT XWK PWEWIWRD XM
 NTJCTYZL OAS XYQ ARVVRKFONT BH SFJDUUXFP
 OUVIGJPF ULBFZ RV DKUKW ROHROZ

IDIOMORPHIC SEQUENCES:

- (1) PWEWIWRD (2) ARVVRKFONT (3) SFJDUUXFP
 (4) ROHROZ

c. Reference to lists of words commonly found in military text and arranged according to their idiomorphic patterns or formulae soon gives suggestions for these cipher groups. Thus:

- (1) PWEWIWRD
DIVISION
 (2) ARVVRKFONT
BATTALIONS
 (3) SFJDUUXFP
ARTILLERY
 (4) ROHROZ
OCLOCK

d. With these assumed equivalents a reconstruction skeleton or diagram of cipher alphabets (forming a portion of a quadricular table) is established, on the hypothesis that the cipher alphabets have been derived from the sliding of a mixed component against the normal sequence. First it is noted that since $O_p = R_c$ both in the word DIVISION and in the word OCLOCK, their cipher equivalents must be in the same alphabet. The reconstruction skeleton is then as follows:

	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	
DIVISION, OCLOCK (1)			O	P					W	Z	H			D	R				I			E					
BATTALION (2)	R	A							F		K			N	O				T	V							
ARTILLERY (3)	S			X					D		U							F	J						P		

Figure 3a.

e. Noting that the interval between O and R in the first and second alphabets is the same, direct symmetry of position is assumed. In a few moments the first alphabet in the skeleton becomes as follows:

	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)		N	O	P		S	T	V	W	X	Z	H		D	R	A	U		I			E	F		J	K
(2)	R	A						F			K		N	O					T	V						
(3)	S			X				D			U							F	J					P		

Figure 3b.

f. The keyword upon which the mixed component is based is now not difficult to find: HYDRAULIC.

g. (1) To decipher the entire message, the simplest procedure is to convert the cipher letters into their plain-component equivalents (setting the HYDRAULIC ...Z sequence against the normal alphabet at any point of coincidence) and then completing the plain-component sequence, as usual. The words of the message will then reappear on different generatrices. The key-letters may then be ascertained and the solution completed. Thus, for the first three words, the diagram 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 ... H Y D R A U L I C B E F G J K M N O P Q S T V U X Z

<u>X I X L P</u>	<u>E Q V I B</u>	<u>V E F H A P F V T</u>
Y H Y G S	K T W H J	W K L A E S L W V
Z I Z H T	L U X I K	X L M B F T M X W
A J A I U	M V Y J L	Y M N C G U N Y X
B K B J V	N W Z K M	Z N O D H V O Z Y
C L C K W	O X A L N	A O P E I W P A Z
D M D L X	P Y B M O	B P Q F J X Q B A
<u>E N E M Y</u>	<u>Q Z C N P</u>	<u>C Q R G K Y R C B</u>
$A_p = S_c$	R A D O Q	D R S H L Z S D C
	S B E P R	<u>E S T I M A T E D</u>
	T C F Q S	$A_p = P_c$
	U D G R T	
	V E H S U	
	W F I T V	
	X G J U W	
	Y H K V X	
	Z I L W Y	
	A J M X Z	
	B K N Y A	
	C L O Z B	
	D M P A C	
	E N Q B D	
	<u>F O R C E</u>	
	$A_p = U_c$	

Figure 4.

(2) The key for the message is found to be SUPREME COURT and the complete message is as follows:

SOLUTION

S U P R E M E C O U R
 ENEMY FORCE ESTIMATED AS ONE DIVISION OF INFANTRY AND TWO BATTALIONS
 XIXLP EQVIB VEFHAPFVT RT XWK PWEIWRD XME NTCCTYZL OAS XYQ ARVVRKFONT

T S U P R E M
 OF ARTILLERY MARCHING NORTH AT SEVEN OCLOCK
 BH SFJDUUXFP OUVIGJPF ULBFZ RV DKUKW ROHROZ

g. In case the plain-component is the reversed normal sequence, the procedure is no different from the foregoing, except that in the completion diagram the reversed sequence is employed after the cipher letters have been converted into their plain-component equivalents.

h. No doubt the student realizes from his previous work that once the primary mixed component has been recovered the latter becomes a known sequence and that the solution of subsequent messages employing the same set of derived alphabets, even though the keys to individual messages are different, then becomes a simple matter.

11. Solution when other types of alphabets are employed. - a. The foregoing examples involve the use either of standard cipher alphabets or of mixed cipher alphabets produced by the sliding of a mixed component against the normal sequence. There is, however, nothing about the general cryptographic scheme which prevents the use of other types of derived, interrelated, or secondary mixed alphabets. Cipher alphabets produced by the sliding of a mixed component against itself (either direct or reversed) or by the sliding of two different mixed components are very commonly encountered in these cases.

b. The solution of such cases involves only slight modifications in procedure, namely, those connected with the reconstruction of the primary components. The student should be in a position to employ to good advantage and without difficulty what he has learned about the principles of indirect symmetry of position in the solution of cases of the kind described.

c. The solution of a message prepared with mixed alphabets derived as indicated in Subpar. b., may be a difficult matter, depending upon the length of the message in question. It might, of course, be almost impossible if the message is short, and there is no background for the application of the probable-word method. But if the message is quite long, or, what is more probable with respect to military communications, should the system be used for regular traffic, so that there are available for study several

messages enciphered by the same set of alphabets, then the problem becomes much easier. In addition to the usual steps in solution by the probable-word method, guided by a search for and identification of idiomorphs, there is the help that can be obtained from the use of the phenomena of isomorphism, a study of which forms the subject of discussion in the next paragraph.

12. Isomorphism and its importance in cryptanalytics. - a. The term idiomorphism is familiar to the student. It designates the phenomena arising from the presence and positions of repeated letters in plain-text words, as a result of which such words may be classified according to their compositions, "patterns", or formulae. The term isomorphism (from the Greek "isos" meaning "equal" and "morphe" meaning "form") designates the phenomena arising from the existence of two or more idiomorphs with identical formulae. Two or more sequences which possess identical formulae are said to be isomorphic.

b. Isomorphism may exist in plain text or in cipher text. For example, the three words WARRANT, LETTERS, and MISSION are isomorphic. If enciphered monoalphabetically, their cipher equivalents would also be isomorphic. In general, isomorphism is a phenomenon of monoalphabeticity (either plain or cipher); but there are instances wherein it is latent and can be made patent in polyalphabetic cryptograms.

c. In practical cryptanalysis the phenomena of isomorphism afford a constantly astonishing source of clues and aids in solution. The alert cryptanalyst is always on the lookout for situations in which he can take advantage of these phenomena, for they are among the most interesting and most important in cryptanalytics.

13. Illustration of the use of isomorphism. - a. Let us consider the case discussed under Par. 10, wherein a message was enciphered with a set of mixed cipher alphabets derived from sliding the keyword-mixed primary

component HYDRAULIC ... XZ against the normal sequence. Suppose the message to be as follows (for simplicity, original word lengths are retained):

CRYPTOGRAM

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

b. (1) Only a few minutes inspection discloses the following three sets of isomorphs:

(1) {	(a) <u>V C L L K I D V S J D C I</u>	(2) {	(a) <u>I X H M P P F X U</u>
	(b) <u>C S P P H Q K C L Z K S Q</u>		(b) <u>H O Q C F F A O X</u>
	(c) <u>P A D D E R X P N B X A R</u>		(3) {
			(b) <u>R O Y X A N O</u>

(2) Without stopping to refer to word-pattern lists in an attempt to identify the very striking idiomorphs of the first set, let the student proceed to build up partial sequences of equivalents, as though he were dealing with a case of indirect symmetry of position. Thus:¹

From isomorphs (1)(a) and (1)(b):

V ⇔ C; C ⇔ S; L ⇔ P; K ⇔ H; I ⇔ Q; D ⇔ K; S ⇔ L; J ⇔ Z

from which the following partial sequences are constructed:

(a) VCSIP (b) DKH (c) IQ (d) JZ

From isomorphs (1)(b) and (1)(c):

C ⇔ P; S ⇔ A; P ⇔ D; H ⇔ E; Q ⇔ R; K ⇔ X; L ⇔ N; Z ⇔ B;

from which the following partial sequences are constructed:

(e) CPD (f) SA (g) HE (h) QR (i) KX (j) LN (k) ZB

From isomorphs (1)(a) and (1)(c):

V ⇔ P; C ⇔ A; L ⇔ D; K ⇔ E; I ⇔ R; D ⇔ X; S ⇔ N; J ⇔ B

¹The symbol ⇔ is to be read "is equivalent to."

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

Figure 7.

b. If, as a result of the analysis of several messages (as described in Par. 25), mixed primary components have been reconstructed, the solution of subsequent messages may readily be accomplished by following the procedure outlined in a above, since in that case the cipher alphabets have become known alphabets.

25. Solution of cipher-text, auto-keyed cryptograms by frequency analysis. - a. Take the short message given in Par. 23c(1). It happens that the letter R_c occurs four times, and because of the nature of the system it also appears four times in the key. It is clear that all plain-text letters enciphered by R_k must be in the same cipher alphabet; in other words, every cipher letter which immediately follows an R_c in the cryptogram belongs to what may be designated conveniently as the R cipher alphabet. The same is

true of every cipher letter following an A_c, a B_c, a C_c, and so on. In short, if a special kind of distribution is made of the text of a cryptogram enciphered in this way, the text can be allocated into 26 uniliteral, monoalphabetic frequency distributions, and the latter can then be solved by frequency analysis, providing there are sufficient data for the purpose in one or more of the distributions.

b. An example will serve to clarify the procedure. Suppose the following cryptograms are at hand:

(1) USYPW TRXDI MLEXR KVDDB DQGSU NSFBO
 BEKVB MAMMO TXXBW ENAXM QLZIX DIXGZ
 PMYUC NEVVJ LKZEK URCNI FQFNN YGSIJ
 TCVNI XDDQQ EKKLR VRFRF XROCS SJTBV
 EFAAG ZRLFD NDSCD MPBBV DEWRR NQICH
 ATNNB OUPIT JLXTC VAOVE YJJLK DMLEG
 NXQWH UVEVY PLQGW UPVKU BMMLB OAEOT
 TNKKU XLODL WTHCZ R

(2) BIWBF GRXLG HOUZO LLZNA MHCTY SCAAT
 XRSCT KVBWK OTGUQ QFJOC YYBVK IXDMT
 KTTCE KVKRO BOEPL QIGNR IQOVJ YKIPH
 JOEYM RPEEW HOTJO CRIIX OZETZ NK

(3) HALOZ JRRVM MHCVB YUHAO EOvac QVVJL
 KZEKU RFRFX YBHALL ZOFHM RSJYL APGRS
 XAGXD MCUNX XLXGZ JPWUI FDBBY PVFZN
 BJNNB ITMLJ OOSEA ATKPB Y

A set of 26 distributions is now made, corresponding to the 26 letters of the alphabet. In each distribution a tally is entered in the appropriate cell to indicate the cipher letter which immediately follows each occurrence of the letter to which the distribution applies.

Key X K Y R Z E C S M M D W A R D D V O S . . .
 Plain N O T I F Y Q U A R T E R M A S T E R . . .
 Cipher ... K Y R Z E C S M M D W A R D D V O S J . . .

(2) If a keyword is used:

Key T Y P E W R I T E R G M I M B P Y N E I . . .
 Plain N O T I F Y Q U A R T E R M A S T E R . . .
 Cipher ... G M I M B P Y N E I Z Q Z Y B H R R V . . .

(3) Sometimes only the last cipher letter resulting from the use of the prearranged keyword is used as the keyletter for enciphering the auto-keyed portion of the text. Thus, in the last example, the plain text beginning TERMASTER would be enciphered as follows:

Key T Y P E W R I T E R I B F W I I A T X . . .
 Plain N O T I F Y Q U A R T E R M A S T E R . . .
 Cipher ... G M I M B P Y N E I B F W I I A T X O . . .

d. In the foregoing examples, direct standard alphabets are employed; but mixed alphabets, either interrelated or independent, may be used just as readily. Also, instead of the ordinary type of cipher alphabets, one may employ a mathematical process of addition (see Par. 40 of Special Text No. 166, Advanced Military Cryptography) but the difference between the latter process and the ordinary one using sliding alphabets is more apparent than real.

e. Since the analysis of the case in which the cipher text constitutes the auto key is usually easier than that in which the plain text serves this function, the former will be the first to be discussed.

24. Solution of cipher-text auto-keyed cryptograms when known alphabets are employed. - a. (1) First of all it is to be noted that if the cryptanalyst knows the cipher alphabets which were employed in encipherment, the solution presents hardly any problem at all. It is only necessary to decipher the message beyond the keyletter or keyword portion and the initial part of the plain text enciphered by this keyletter or keyword can be filled

in from the context. An example, using standard cipher alphabets, follows herewith:

Cipher ... W S G Q V O H V M Q W E Q U H A A L N B N Z Z M P E S K D

(2) Writing the cipher text as keyletters (displaced one interval to the right) and deciphering by direct standard alphabets yields the following:

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

(3) Trial of the word REPORT as the initial word of the message yields an intelligible word as the initial key: FORCE, so that the message reads:

Key	F O R C E V O H V M Q
Cipher ...	W S G Q V O H V M Q
Plain	R E P O R T T O R E

(4) A semi-automatic method of solving such a message is to use sliding normal alphabets and align the strips so that, as one progresses from left to right, each cipher letter is set opposite the letter A on the preceding strip. Taking the letters VMQWEQUH in the foregoing example, note the following series of placements of the successive strips. Then note how the successive plain-text letters of the word REGIMENT reappear to the left of the successive cipher letters MQWEQUH.

advance; if the interruptor is a plain-text letter, while the interruptions can be indicated before encipherment is begun, the irregularities occasioned by the interruptions in keying cause confusion and quite materially retard the enciphering process. In deciphering, the rate of speed would be just as slow in either method. It is obvious that one of the principal disadvantages in all these methods is that if an error in transmission is made, if some letters are omitted, or if anything happens to the interruptor letter, the message becomes difficult or impossible to decryptograph by the ordinary code clerk. Finally, the degree of cryptographic security attainable by most of these methods is not sufficient for military purposes.

SECTION VI.

SOLUTION OF AUTO-KEY SYSTEMS, I.

	Paragraph
The two basic methods of auto-key encipherment	23
Solution of cipher-text auto-keyed cryptograms when known alphabets are employed	24
Solution of cipher-text auto-keyed cryptograms by frequency analysis	25
Special case of solution of cipher-text auto-keyed cryptograms	26

23. The two basic methods of auto-key encipherment. - a. In auto-key encipherment there are two possible sources for successive keyletters: the plain text or the cipher text of the message itself. In either case, the initial keyletter or keyletters are supplied by preagreement between the correspondents; after that the text letters that are to serve as the key are displaced 1, 2, 3, . . . intervals to the right, depending upon the length of the prearranged key.

b. (1) An example of plain-text keying will first be shown, to refresh the student's recollection. Let the previously-agreed upon key consist of a single letter, say X, and let the cipher alphabets be direct standard alphabets.

```
Key ..... X N O T I F Y Q U A R T E R M A S T E R . . .
Plain .... N O T I F Y Q U A R T E R M A S T E R . . .
Cipher ... K B H B N D O K U R K X V D M S L X V
```

(2) Instead of having a single letter serve as the initial key, a word or even a long phrase may be used. Thus (using TYPEWRITER as the initial key):

```
Key ..... T Y P E W R I T E R | N O T I F Y Q U A R . . .
Plain .... N O T I F Y Q U A R T E R M A S T E R . . .
Cipher ... G M I M B P Y N E I G S K U F Q J Y R . . .
```

c. (1) In cipher-text auto-keying the procedure is quite similar.

If a single initial keyletter is used:

- 43 -

counterbalanced by the fact that whereas in the former case the cryptanalyst is dealing with the initial words of messages, in this case he is dealing with interior portions of the text and has no way of knowing where a word begins. The latter remarks naturally do not apply to the case where a whole set of messages in this system, all in the same key, can be subjected to simultaneous study. In such a case the cryptanalyst would also have the initial words to work upon.

22. Concluding remarks. - a. The preceding two paragraphs both deal with the first and simplest of the three basic cases referred to under Par. 18. The second of those cases involves considerably more work in solution for the reason that when the interruption takes place and the keying sequence recommences, the latter is not invariably the initial point of the sequence, as in the first case.

b. In the second of those cases the interruptor causes a break in the keying sequence and a recommencement at any one of the 10 keying elements. Consequently, it is impossible now merely to superimpose sections of the text by shifting them so that their initial letters fall in the same column. But a superimposition is nevertheless possible, provided the interruptions do not occur so frequently that sections of only a very few letters are enciphered by sequent keyletters. In order to accomplish a proper superimposition in this case, a statistical test would be essential, and for this a good many letters are required. The nature of this test will be explained in Section X.

c. The same thing is true of the last of the three cases mentioned under Par. 18. The solution of a case of this sort is admittedly a rather difficult matter which will be taken up in its proper place later.

d. (1) In the cases thus far studied, either the plain-text groupings were variable in length and were enciphered by a constant-length key, or the plain-text groupings were constant in length and were enciphered by a variable-length key. It is possible, however, to combine both principles and to apply a variable-length key to variable-length groupings of the plain text.

(2) Suppose the communicants agree to encipher a message according to word lengths but, at irregular intervals, to add at the end of a word an interruptor letter which will serve to interrupt the key. Note the following, in which the key is BUSINESS MACHINES and the interruptor letter is X:

Key.....	B	U	S	B	
Plain....	A M M U N I T I O N	F O R	F I R S T X	A R T I L L E R Y	etc.
Cipher...	B T T R V O D O W V	E Q V	Z D F G J O	B H D O S S J H I	

Cryptogram:

B T T R V O D O W V E Q V Z D F G J O B H D O S S J H I... etc.

(3) The foregoing system is only a minor modification of the simple case of ordinary word length encipherment as explained in Sect. II. If standard cipher alphabets are used, the spasmodic interruption and the presence of the interruptor letter would cause no difficulty whatever, since the solution can be achieved mechanically, by completing the plain-component sequence. If unknown mixed cipher alphabets are used, and the primary components are unknown, solution may be reached by following the procedure outlined in Sections II and III, with such modifications as are suitable to the case.

e. It is hardly necessary to point out that the foregoing types of aperiodic substitution are rather unsuitable for practical military usage. Encipherment is slow and subject to error. In some cases encipherment can be accomplished only by single-letter operation. For if the interruptor is a cipher letter the key is interrupted by a letter which cannot be known in

two distributions, frequency studies being aided by considerations based upon probable words. In this case, since the text comprises only the beginnings of messages, assumptions for probable words are more easily made than when words are sought in the interiors of messages. Such common introductory words as REQUEST, REFER, ENEMY, WHAT, WHEN, IN, SEND, etc., are good ones to assume. Furthermore, high-frequency digraphs used as the initial digraphs of common words will of course manifest themselves in the first two columns. The greatest aid in this process is, as usual, a familiarity with the "word habits" of the enemy.

(4) Let the student try to solve the messages. In so doing he will more or less quickly find the cause of the rapid falling off in mono-alphabeticity as the columns progress to the right from the initial point of the messages.

21. Interruptor is a cipher-text letter. - a. In the preceding case a plain-text letter serves as the interruptor. But now suppose the communicants agree that the interruption in the key will take place immediately after a previously-agreed-upon letter, say Q, occurs in the cipher text. The key would then be interrupted as shown in the following example:

Key.....	BUSINESSMACHINESBUSINESSM	
Plain....	AMMUNITIONFORFIRSTARTILLE	
Cipher...	BOLYRPJDROJKXTPFYXSXBPUUQ	

Key.....	BUSINESSMACHIN	BUSINESSMACH	BU
Plain....	RYWILLBELOADED	AFTERAMMUNIT	IO
Cipher...	HRNMYTTXHPCRFQ	BEJFIELLBONQ	OOQ

Key.....	BUSINESSMACH	BUSINE
Plain....	NFORTHIRDART	ILLERY
Cipher...	VECXBODFPAZQ	ONUFIC

Cryptogram:

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

- 42 -

b. In the foregoing example, there are no significant repetitions. Such as do occur comprise only digraphs, one of which is purely accidental. But the absence of significant, long repetitions is itself purely accidental; for had the interruptor letter been a letter other than Q_c , then the phrase AMMUNITION FOR, which occurs twice, might have been enciphered identically both times. If a short key is employed, repetitions may be plentiful. For example, note the following, in which S_c is the interruptor letter:

Key.....	BANDSBANDSBANDSBANDSBAN		BANDSBANDSB
Plain....	FROMFOURFIVETOFOURFIFTE		ENAMBARRAGE
Cipher...	<u>K T A K Z W X I I D A C B N</u>		<u>Z W X I I D K W S J O N K T B T I D H J</u>

c. This last example gives a clue to one method of attacking this type of system. There will be repetitions within short sections, and the interval between them will sometimes permit of ascertaining the length of the key. In such short sections, the letters which intervene between the repeated sequences may be eliminated as possible interruptor letters. Thus, the letters A, C, B, and N may be eliminated, in the foregoing example, as interruptor letters. By extension of this principle to the letters intervening between other repetitions, one may more or less quickly ascertain what letter serves as the interruptor.

d. Once the interruptor letter has been found, the next step is to break up the message into "uninterrupted" sequences and then attempt a solution by superimposition. The principles explained in Par. 20 need only be modified in minor respects. In the first place, in this case the columns of text formed by the superimposition of uninterrupted sequences will be purely monoalphabetic, whereas in the case of the example in Par. 20, only the very first column is purely monoalphabetic, the monoalphabeticity falling off very rapidly with the 2d, 3d, ... columns. Hence, in this case the analysis of the individual alphabets should be an easier task. But this would be

imposition can be employed, provided the messages can be superimposed correctly, that is, so that the letters which fall in one column really belong to one cipher alphabet. Just how this can be done will be demonstrated in subsequent paragraphs, and a clue has already been given in Par. 18c. At this point, however, a simple illustration of the method will be given, using the substitution system discussed in Par. 19.

b. Example. (1) A set of 35 messages has been intercepted on the same day. Presumably they are all in the same key, and the presence of repetitions between messages corroborates this assumption. But the intervals between repetitions within the same message do not show any common factor and the messages appear to be aperiodic in nature. The probable-word method has been applied, using standard alphabets, with no success. The messages are then superimposed (Fig. 5); the frequency distributions for the first 10 columns are as shown in Fig. 6.

1	ZCTPZWPZPEPZQX	19	AFEOJTDTIT
2	WTEQMXYZSPRC	20	KPVFQWPKTEV
3	TCRWCXTBHH	21	ZABGRTPXPUQX
4	EFKCSZRIHA	22	YHEOCUHM DT
5	YANCIHZNUW	23	CLCPZIKOTH
6	VZIETIRRGX	24	AFLWWZQM DT
7	HCQICKGUON	25	ZCWAPMBSAWL
8	ZCFCLXRKQW	26	HFLMHRZNAPECEI
9	HWWPTEWCIMJS	27	CLZGEMKZTO
10	EPDOZCLIKSJ	28	TPYFKOTIZUH
11	WTSSQZPZIET	29	ZCCPSNEOPHDYL
12	ZCGGYFC SBG	30	CIYGIFTSYTLE
13	CWZAOOEMHWTP	31	YTSVWVDGHPGUZ
14	CIYGIFBDTVX	32	NOCAIFBJBLGHY
15	EAQDRDNSRCAPDT	33	ZXXFLFEGJL
16	YFWCQQBZCWC	34	ZCTMMBZJOO
17	WTEZQSKUHC	35	HCQIWSYSBPHCZV
18	ZCVXQZKZYDWLK		

Figure 5

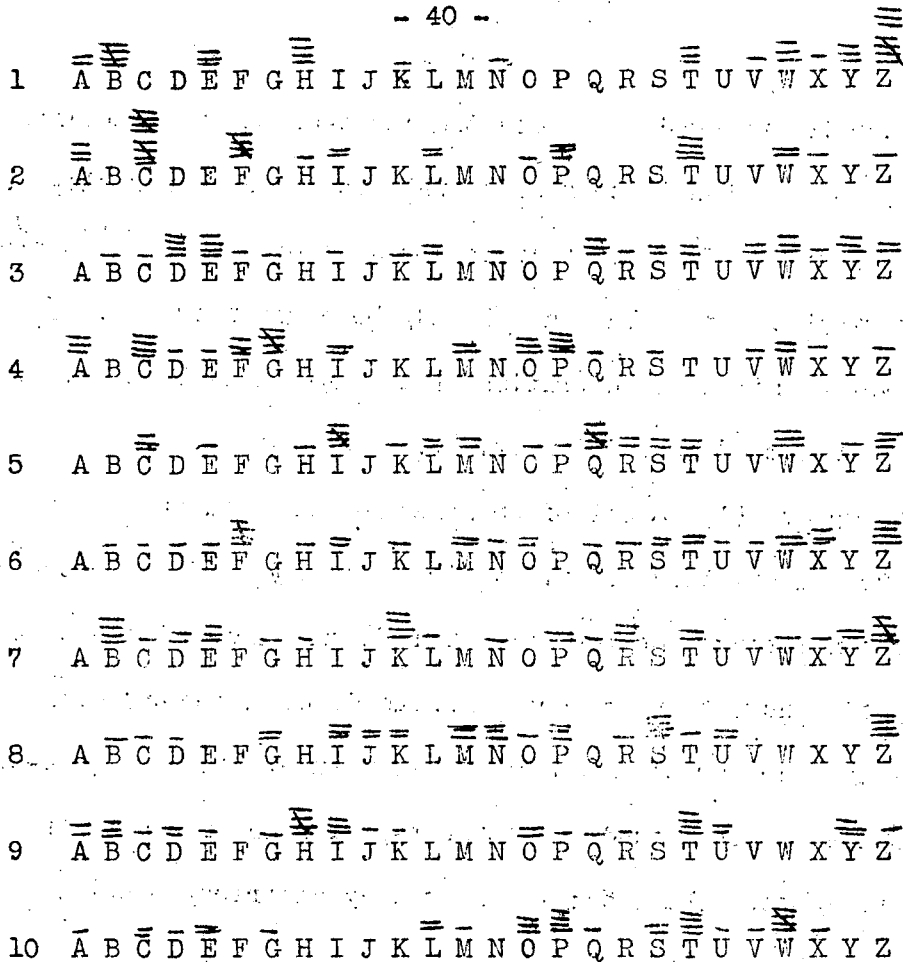


Figure 6.

(2) The 1st and 2d distributions are certainly monoalphabetic.

There are very marked crests and troughs, and the number of blanks (15 and 14, respectively) is more than satisfactory in both cases. (Let the student at this point refer to Par. 14 and Chart 5 of Military Cryptanalysis, Part I.) But the 3d, 4th, and remaining distributions appear no longer to be monoalphabetic. Note particularly the distribution for the 6th column. From this fact the conclusion is drawn that some disturbance in periodicity has been introduced in the cryptograms. In other words, although they all start out with the same alphabet, some sort of interruption takes place so as to suppress periodicity.

(3) However, a start on solution may be made by attacking the first

(2) Since this "key" is certainly not intelligible text, the assumed word is moved one letter to the right and the test repeated, and so on until the following place in the test is reached:

Cipher... S X D J P U S Y I
 Plain.... A R T I L L E R Y
 Key..... S I B U S I N E B

(3) The sequence BUSINE suggests BUSINESS; moreover, it is noted that the key is interrupted both times by the letter R_p. Now the key may be applied to the beginning of the message, to see if the whole key or only a portion of it has been recovered. Thus:

Key..... B U S I N E S S B U S
 Cipher... B O L Y R P J D R O J
 Plain.... A M M U N I T I U M T

(4) It is obvious that BUSINESS is only a part of the key. But the deciphered sequence certainly seems to be the word AMMUNITION. When this is tried, the key is extended to BUSINESS MA... . Enough has been shown to clarify the procedure.

f. The foregoing solution is predicated upon the hypothesis that the cipher alphabets are known. But what if this is not the case? What of the steps necessary to arrive at the first solution, before even the presence of an interruptor is suspected? The answer to this question leads to the presentation of a method of attack which is one of the most important and powerful means the cryptanalyst has at his command for many knotty problems. It is called solution by superimposition, and warrants detailed treatment.

20. Solution by superimposition. - a. Basic principles. - (1) In solving an ordinary repeating-key cipher the first step, that of ascertaining the length of the period, is of no significance in itself. It merely paves the way for and makes possible the second step, which consists in allocating the letters of the cryptogram into individual monoalphabetic distributions.

The third step then consists in solving these distributions. Usually, the text of the message is transcribed into its periods and is written out in successive lines corresponding in length with that of the period. The diagram then consists of a series of columns of letters and the letters in each column belong to the same monoalphabet. Another way of looking at the matter is to conceive of the text as having thus been transcribed into superimposed periods; the letters in each column have undergone the same kind of treatment by the same elements (plain and cipher components of the cipher alphabet).

(2) Suppose, however, that the repetitive key is very long and that the message is short, so that there are only a very few cycles in the text. Then the solution of the message becomes difficult, if not impossible, because there is not a sufficient number of superimposable periods to yield monoalphabetic distributions which can be solved by frequency principles. But suppose also that there are many short cryptograms all enciphered by the same key. Then it is clear that if these messages are superimposed:

(a) The letters in the respective columns will all belong to individual alphabets; and

(b) If there is a sufficient number of such superimposable messages (say 25 - 30, for English), then the frequency distributions applicable to the successive columns of text can be solved -- without knowing the length of the key. In other words, any difficulties that may have arisen on account of failure or inability to ascertain the length of the period have been circumvented. The second step in normal solution is thus "by-passed".

(3) Furthermore, and this is a very important point, in case an extremely long key is employed and a series of messages beginning at different initial points are enciphered by such a key, this method of solution by super-

Cryptogram:

B O L Y R P J D R O J K X K J F Y X S X D J U P S Y I Y D P Y F
X U R A F A E N M J J V B O L Y R P J D R O J K X D G D X G U F D
J U P S Y I X X X X

b. Instead of employing an ordinary plain-text letter as the interruptor, one might reserve the letter J for this purpose (and use the letter I whenever this letter appears as part of a plain-text word). This is a quite simple variation of the basic method. The letter J acts merely as though it were a plain-text letter, except that in this case it also serves as the interruptor. The interruptor is then inserted at random, at the whim of the enciphering clerk. Thus:

Key.....	BUSINESSMAC	BUSINESSM	BUSINESSMACHINESBUSIN
Plain...	TROOPSWILLJ	BEHALTEDJ	ATROADIUNCTIONFIVESIX

c. It is obvious that repetitions would be plentiful in cryptograms of this construction, regardless of whether a letter of high, medium, or low frequency is selected as the signal for key interruption. If a letter of high frequency is chosen, repetitions will occur quite often, not only because that letter will certainly be a part of many common words, but also because it will be followed by words that are frequently repeated; and since the key starts again with each such interruption, these frequently-repeated words will be enciphered by the same sequence of alphabets. This is the case in the first of the two foregoing examples. It is clear, for instance, that every time the word ARTILLERY appears in the cryptogram the cipher equivalents of TILLERY must be the same. If the interruptor letter were A_p instead of R_p , the repetition would include the cipher equivalents of RTILLERY; if it were T_p , ILLERY, and so on. On the other hand, if a letter of low frequency were selected as the interruptor letter, then the encipherment would tend to approximate that of normal repeating-key substitution, and repetitions would be

plentiful on that basis alone.

d. Of course, the lengths of the intervals between the repetitions, in any of the foregoing cases, would be irregular, so that periodicity would not be manifested. The student may inquire, therefore, how one would proceed to solve such messages, for it is obvious that an attempt to allocate the letters of a single message into separate monoalphabetic distributions cannot be successful unless the exact locations of the interruptions are known -- and they do not become known to the cryptanalyst until he has solved the message, or at least a part of it. Thus it would appear as though the would-be solver is here confronted with a more or less insoluble dilemma. This sort of reasoning, however, makes more of an appeal to the uninitiated observer than to the experienced cryptanalyst, who specializes in methods of solving cryptographic dilemmas.

e. (1) The problem here will be attacked upon the usual two hypotheses, and the easier one will be discussed first. Suppose the system has been in use for some time, that an original solution has been reached by means to be discussed under the second hypothesis, and that the cipher alphabets are known. There remains unknown only the specific key to messages. Examining whatever repetitions are found, an attack is made on the basis of searching for a probable word. Thus, taking the illustrative message in subpar. a, suppose the presence of the word ARTILLERY is suspected. Attempts are made to locate this word, basing the search upon the construction of an intelligible key. Beginning with the very first letter of the message, the word ARTILLERY is juxtaposed against the cipher text, and the key letters ascertained, using the known alphabets, which we will assume in this case are based upon the HYDRAULIC...XZ sequence sliding against the normal. Thus:

Cipher...	B O L Y R P J D R
Plain...	A R T I L L E R Y
"Key"...	B H J Q P I B F U

the matter may become quite complex.

c. If one knows when the interruptions take place in each cycle, then successive sections of the basic keying cycle in the three cases may be superimposed. Thus:

		Method (1)									
		KEYING ELEMENT NUMBER									
		1	2	3	4	5	6	7	8	9	10
Letter No.		1	2	3	4						
Letter No.		5	6	7	8	9	10				
Letter No.		11	12	13							
Letter No.		14	15	16	17	18	19	20			
Letter No.		21	22	23	24	25	26	27	28	29	30
Letter No.		31	32	33							
Letter No.		34	35	etc.							

		Method (2)									
		KEYING ELEMENT NUMBER									
		1	2	3	4	5	6	7	8	9	10
Letter No.		1	2	3	-	-	-	4	5	6	7
Letter No.		8	9	-	10	11	12				
Letter No.		-	-	13	14	15	16	17	18	19	20
Letter No.		21	-	-	-	-	-	-	22	23	24
Letter No.		25	26	-	-	27	28	29	-	30	31
Letter No.		32	-	-	-	33	34	35	etc.		

		Method (3)									
		KEYING ELEMENT NUMBER									
		1	2	3	4	5	6	7	8	9	10
Letter No.		1	2	3	4	5	-	-	-	-	-
Letter No.		-	-	7	6	-	-	-	-	-	-
Letter No.		-	-	-	8	9	10	11	12	13	14
Letter No.		15	-	-	-	-	-	19	18	17	16
Letter No.		23	24	25	-	-	-	-	20	21	22
Letter No.		27	26	-	-	-	-	-	30	29	28
Letter No.		33	34	35	etc.					31	32

- 34 -

or how

Obviously if one does not know when/the interruptions take place, then the successive sections of keying elements can not be superimposed as indicated above.

d. The interruption of the cyclic keying sequence usually takes place according to some prearranged plan, and the three basic methods of interruption will be taken up in turn, using a short mnemonic key as an example.

e. Suppose the correspondents agree that the interruption in the keying sequence will take place after the occurrence of a specified letter called an interruptor¹, which may be a letter of the plain text, or one of the cipher text, as agreed upon in advance. Then, since in either case there is nothing fixed about the time the interruption will occur -- it will take place at no fixed intervals -- not only does the interruption become quite irregular, following no pattern, but also the method never reverts to one having periodicity. Methods of this type will now be discussed in detail.

19. Interruptor is a plain-text letter. - a. Suppose the communicants agree that the interruption in the key will take place immediately after a previously agreed-upon letter, say R, occurs in the plain text. The key would then be interrupted as shown in the following example (using the mnemonic key BUSINESS MACHINES and the HYDRAULIC ... XZ sequence):

Key.....	BUSINESSMACHI	BUS	BUSI	BUSINE
Plain....	AMMUNITIONFOR	FIR	STAR	TILLER
Cipher...	BOLYRPJDROJKX	KJF	YXSX	DJUPSY

Key.....	BUSINESSMACHINESBU	BUSINESSMACHI
Plain....	YWILLBELOADEDAFTER	AMMUNITIONFOR
Cipher...	IYDPYFXURAFENMJJV	BOLYRPJDROJKX

Key.....	BUSI	BUS	BUSINE	BUSIN
Plain....	THIR	DAR	TILLER	Y
Cipher...	DGD X	GUF	DJUPSY	I

¹ Also called at times an "influence" letter because it influences or modifies normal procedure.

SECTION V

SOLUTION OF SYSTEMS USING VARIABLE-LENGTH KEYING
UNITS TO ENCIPHER CONSTANT-LENGTH PLAIN-TEXT GROUPINGS

	Paragraph
Variable-length groupings of the keying sequence	17
Methods of interrupting a cyclic keying sequence	18
Interruptor is a plain-text letter	19
Solution by superimposition	20
Interruptor is a cipher-text letter	21
Concluding remarks	22

17. Variable-length groupings of the keying sequence. - The preceding cases deal with simple methods of eliminating or avoiding periodicity by enciphering variable-length groupings of the plain text, using constant-length keying units. In Par. 2a, however, it was pointed out that periodicity can also be suppressed by applying variable-length key groupings to constant-length plain-text groups. One such method consists in irregularly interrupting the keying sequence, if the latter is of a limited or fixed length, and recommencing it (from its initial point) after such interruption, so that the keying sequence becomes equivalent to a series of keys of different lengths. Thus, the keyphrase BUSINESS MACHINES may be expanded to a series of irregular-length keying sequences, such as BUSI/BUSINE/BU/BUSINESSM/BUSINESSMAC, etc. Various schemes or prearrangements for indicating or determining the interruptions may be adopted. Three methods will be mentioned in the next paragraph.

18. Methods of interrupting a cyclic keying sequence. - a. There are many methods of interrupting a keying sequence which is basically cyclic, and which therefore would give rise to periodicity if not interfered with in some way. These methods may, however, be classified into three categories as regards what happens after the interruption occurs:

- (1) The keying sequence merely stops and begins again at the initial

point of the cycle.

(2) One or more of the elements in the keying sequence may be omitted from time to time irregularly.

(3) The keying sequence irregularly alternates in its direction of progression, with or without omission of some of its elements.

b. These methods may, for clarity, be represented graphically as follows. Suppose the key consists of a cyclic sequence of 10 elements represented symbolically by the series of numbers 1, 2, 3, ..., 10. Using an asterisk to indicate an interruption, the following may then represent the relation between the letter number of the message and the element number of the keying sequences in the three types mentioned above.

(1)
Letter No..... 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Key element No... 1-2-3-4-*1-2-3-4-5- 6-*1- 2- 3-*1- 2- 3- 4- 5- 6- 7-*

Letter No..... 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Key element No... 1- 2- 3- 4- 5- 6- 7- 8- 9-10- 1- 2- 3-*1- 2- etc.

(2)
Letter No..... 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Key element No... 1-2-3-*7-8-9-10-1-2-*4- 5- 6-*3- 4- 5- 6- 7- 8- 9-10-

Letter No..... 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Key element No... 1-*8- 9-10- 1- 2-*5- 6- 7-*9-10- 1-*5- 6- 7- etc.

(3)
Letter No..... 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Key element No... 1-2-3-4-5-*4-3-*4-5- 6- 7- 8- 9-10- 1-*10- 9- 8- 7-*8

Letter No..... 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Key element No... 9-10- 1- 2- 3-*2- 1-10- 9- 8-*9-10- 1- 2- 3 etc.

As regards the third method, which involves only an alternation in the direction of progression of the keying sequence, if there were no interruptions in the key it would mean merely that a 10-element keying sequence, for example, could be treated as though it were an 18-element sequence and the matter could then be handled as though it were a special form of the second method. But if the principles of the second and third method are combined in one system,

occurrences of the word (the sum of the values of the letters of the keyphrase = 140); and then the chances that the keyletter P would begin the encipherment of DIV are but one in three. Only one of these three possible encipherments will yield exactly the same sequence of cipher equivalents the second time as was obtained the first time. For example, if the text were such as to place two occurrences of the word DIVISION in the positions shown below, their encipherments would be as follows:

3	1	1	3	2	1	1	2	3	1	3	1	1	3	2	1	1	2	3	1
P	R	E	P	A	R	E	D			P	R	E	P	A	R	E	D	U	N
FIR	S	T	DIV	IS	I	O	NDI	VI	S	I	ON
...	THJ	GV	F	X	MTH	ZG	T	P	NM

Although the word DIVISION, on its second appearance, begins but one letter beyond the place where it begins on its first appearance, the cipher equivalents now agree only in the first two letters, the fourth, and the last letters. Thus:

	<u>D</u>	<u>I</u>	<u>V</u>	<u>I</u>	<u>S</u>	<u>I</u>	<u>O</u>	<u>N</u>
(1)	<u>T</u>	<u>H</u>	<u>J</u>	<u>G</u>	<u>V</u>	<u>F</u>	<u>X</u>	<u>M</u>
(2)	<u>T</u>	<u>H</u>	<u>Z</u>	<u>G</u>	<u>T</u>	<u>P</u>	<u>N</u>	<u>M</u>

e. Attention is directed to the characteristics of the foregoing two encipherments of the same word. When they are superimposed, the first two cipher equivalents are the same in the two encipherments; then there is a single interval where the cipher equivalents are different; the next cipher equivalent is the same; then follow three intervals with dissimilar cipher equivalents; finally, the last cipher equivalent is the same in both cases. The repetitions here extend only to one or two letters; longer repetitions can occur only exceptionally. The two encipherments yield only occasional coincidences, that is, places where the cipher letters are identical; moreover, the distribution of the coincidences is quite irregular and of an intermittent character.

f. This phenomenon of intermittent coincidences, involving coincidences of single letters, pairs of letters, or short sequences (rarely ever exceeding pentagraphs) is one of the characteristics of this general class of polyalphabetic substitution, wherein the cryptograms commonly manifest what appears to be a disturbed or distorted periodicity.

g. From a technical standpoint, the cryptographic principle upon which the foregoing system is based has much merit, but for practical usage it is entirely too slow and too subject to error. However, if the encipherment were mechanized by machinery, and if the enciphering key were quite lengthy, such a system and mechanism becomes of practical importance. Cipher machines for accomplishing this type of substitution will be treated in a subsequent text.

irregular groupings in encipherment. The determination of the length of the cycle might, however, present difficulties in some cases, since the basic or fundamental period would not be clearly evident because of the presence of repetitions which are not periodic in their origin. For example, suppose the word PREPARE were used as a key, each keyletter being employed to encipher a number of letters corresponding to its numerical value in the normal sequence. It is clear that the length of the basic period, in terms of letters, would here be the sum of the numerical values of P (= 16) + R (= 18) + E (= 5), and so on, totalling 79 letters. But because the key itself contains repeated letters and because encipherment by each keyletter is monoalphabetic, there would be plenty of cases in which the first letter P would encipher the same or part of the same word as the second letter P, producing repetitions in the cryptogram. The same would be true as regards encipherments by the two R's and the two E's in this keyword. Consequently, the basic period of 79 would be distorted or masked by aperiodic repetitions, the intervals between which would not be a function of, nor bear any relation to, the length of the key. The student will encounter more cases of this kind, in which a fundamental periodicity is masked or obscured by the presence of cipher-text repetitions not attributable to the fundamental cycle. The experienced cryptanalyst is on the lookout for phenomena of this type, when he finds in a polyalphabetic cipher plenty of repetitions but with no factorable constancy which leads to the disclosure of a short period. He may conclude, then, either that the cryptogram involves several primary periods which interact to produce a long resultant period, or that it involves a fairly long fundamental cycle within which repetitions of a nonperiodic origin are present and obscure the phenomena manifested by repetitions of a periodic origin.

d. (1) A logical extension of the principle of polyalphabetic encipherment of variable-length plain-text groupings is the case in which these plain-

text groupings rarely exceed 4 letters, so that a given cipher alphabet is in play for only a very short time, thus breaking up what might otherwise appear as fairly long repetitions in the cipher text. For example, suppose the letters of the alphabet, arranged in their normal-frequency order, were set off into four groups, as follows:

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

(2) Suppose that a letter in group 1 means that one letter will be enciphered; a letter in group 2, that two letters will be enciphered; and so on. Suppose, next, that a rather lengthy phrase were used as a key, for example, PREPARED UNDER THE DIRECTION OF THE CHIEF SIGNAL OFFICER FOR USE WITH ARMY EXTENSION COURSES. Suppose, finally, that each letter of the key were used not only to select the particular cipher alphabet to be used, but also to control the number of letters to be enciphered by the selected alphabet, according to the scheme outlined above. Such an enciphering scheme, using the HYDRAULIC...XZ primary cipher component sliding against the normal plain component, would yield the following groupings:

Grouping...	3	1	1	3	2	1	1	2	3	1	2	1	1	1	3	1	2
Key.....	P	R	E	P	A	R	E	D	U	N	D	E	R	T	H	E	D
Plain.....	FIR	S	T	DIV	IS	I	O	NW	ILL	A	DV	A	N	C	EAT	F	IV
Cipher.....	WHB	T	R	<u>THJ</u>	GV	F	X	MX	JNN	N	UW	E	N	W	AHQ	M	EW

Grouping...	1	1	1	2	1	1	2	1	2	3	1	3	1	2	3	1	
Key.....	I	R	E	C	T	I	O	N	O	F	T	H	E	C	H	I	...
Plain.....	E	F	I	FT	E	E	NA	M	AS	SEC	O	NDD	I	VI	SIO	N	...
Cipher.....	F	C	P	JY	Z	F	AO	D	OB	RMJ	B	<u>JRR</u>	P	RN	<u>PCK</u>	S	...

(3) Here it will be seen that any tendency for the formation of lengthy repetitions would be counteracted by the short groupings and quick shifting of alphabets. The first time the word DIVISION occurs it is enciphered as THJGVFXM; the second time it occurs it is enciphered as RPRNPCKS. Before DIVISION can be twice enciphered by exactly the same sequence of key letters, an interval of at least 140 letters must intervene between the two

messages or when there are many short messages, a study of isomorphism will disclose a sufficient number of partial isomorphs to give data usually sufficient for purposes of alphabet reconstruction.

e. It should be noted that there is nothing about the phenomenon of isomorphism which restricts its use to cases in which the cipher alphabets are secondary alphabets resulting from the sliding of a mixed component against the normal. It can be useful in all cases of interrelated secondary alphabets no matter what the basis of their derivation may be.

f. In subsequent studies the important role which the phenomenon of isomorphism plays in cryptanalytics will become more apparent. When the traffic is stereotypic in character, even to a slight degree, so that isomorphism may extend over several words or phrases, the phenomenon becomes of highest importance to the cryptanalyst and it becomes an extremely valuable tool in his hands.

15. Word separators. - a. One of the practical difficulties in employing systems in which the keying process shifts according to word lengths is that in handling such a message the decryptographing clerk is often not certain exactly when the termination of a word has been reached, and thus time is lost by him. For instance, while decryptographing a word such as INFORM the clerk would not know whether he now has the complete word and should shift to the next keyletter or not: the word might be INFORMS, INFORMED, INFORMING, INFORMAL, INFORMATION, etc. The past tense of verbs, the plural of nouns, and terminations of various sorts capable of being added to word roots would give rise to difficulties, and the latter would be especially troublesome if the messages contained a few telegraphic errors. Consequently, a scheme which is often adopted to circumvent this source of trouble is to indicate the end of a word by an infrequent letter such as Q or X, and enciphering the letter; in such usage these letters are called word separators.

b. When word separators are employed and this fact is once discovered, their presence is of as much aid to the cryptanalyst in his solution as it is to the clerks who are to decryptograph the messages. Sometimes the presence of these word separators, even when enciphered, aids or makes possible the blocking out of isomorphs.

16. Variations and concluding remarks on foregoing systems. - a. The systems thus far described are all based upon word-length encipherment using different cipher alphabets. Words are markedly irregular in regard to this feature of their construction, and thus aperiodicity is imparted to such cryptograms. But variations in the method, aimed at making the latter somewhat more secure, are possible. Some of these variations will now be discussed.

b. Instead of enciphering according to natural word lengths, the irregular groupings of the text may be regulated by other agreements. For example, suppose that the numerical value (in the normal sequence) of each keyletter be used to control the number of letters enciphered by the successive cipher alphabets. Depending then upon the composition of the keyword or keyphrase, there would be a varying number of letters enciphered in each alphabet. If the keyword were PREPARE, for instance, then the first cipher alphabet would be used for 16 (P) letters, the second cipher alphabet, for 18 (R) letters, and so on. Monoalphabetic encipherment would therefore allow plenty of opportunity for tell-tale word patterns to manifest themselves in the cipher text. Once an entering wedge is found in this manner, solution would be achieved rather rapidly. Of course, all types of cipher alphabets may be employed in this and the somewhat similar schemes described.

c. If the key is short, and the message is long, periodicity will be manifested in the cryptogram, so that it would be possible to ascertain the length of the basic cycle (in this case the length of the key) despite the

found that many messages begin with the expression REFERRING TO YOUR NUMBER... Having several messages for study, the selection of one which begins with such a common idiomorphism as that given by the word REFERRING is a relatively simple matter; and having found the word REFERRING, if with a fair degree of certainty one can add the words TO YOUR NUMBER, the solution is probably well under way.

c. (1) Take the case discussed in Par. 13, but assume that word lengths are no longer indicated because the message is transmitted in the usual 5-letter groups. The process of ascertaining the exact length of sequences which are isomorphic, or, as the process is briefly termed, "blocking out isomorphs" becomes a more difficult matter and must often rest upon rather tenuous threads of reasoning. For example, take the illustrative message just dealt with and let it be assumed that it was arranged in 5-letter groups:

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

(2) The detection of isomorphisms now becomes a more difficult matter. There is no special trouble in picking out the following three isomorphic sequences:

(1) V C L L K I D V S J D C I

(2) C S P P H Q K C L Z K S Q

(3) P A D D E R X P N B X A R

since the first one happens to be at the beginning of the message and its left-hand boundary, or "head", is marked by (or rather, coincides with) the beginning of the message. By a fortunate circumstance, the right-hand bound-

ary, or "tail", can be fixed just as accurately. That the repetition extends as far as indicated above is certain for we have a check on the last column I, Q, R. If an additional column were added, the letters would be O, L, I. Since the second letter has previously appeared while the first and third have not, a contradiction results and the new column may not be included.

If, however, no one of the three letters O, L, I had previously appeared so that there could be no means of getting a check on their correctness, it would not be possible to block out or ascertain the extent of the isomorphism in such a case. All that could be said would be that it seems to include the first thirteen letters, but it might continue further.

d. (1) However, the difficulty or even the impossibility of blocking out the isomorphs to their full extent is not usually a serious matter. After all, the cryptanalyst uses the phenomenon not to identify words but to obtain cryptanalytic data for reconstructing cipher alphabets. For example, how many data are lost when the illustrative message of subpar. 13a is rewritten in 5-letter groups as in subpar. 14c (1)? Suppose the latter form of message be studied for isomorphs:

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

(2) If the underscored sequences are compared with those in the message in subpar. 13a, it will be found that only a relatively small amount of information has been lost. Certainly not enough to cause any difficulty have been lost in this case, for all the data necessary for the reconstruction of the mixed cipher component came from the first set of isomorphs, and the latter are identical in length in both cases. Only the head and tail letters of the second pair of isomorphic sequences are not included in the underscored sequences in the 5-letter version of the message. The third pair of isomorphic sequences shown in Par. 13b does not appear in the 5-letter version since there is only one repeated letter in this case. In long

- 21 -

(4) From isomorphs (2)(a) and (2)(b), the interval between H and I is 7; it is the same for O and X, Q and H, C and M, etc. From isomorphs (3)(a) and (3)(b) the interval between R and N is 13; it is the same for O and A, Y and K, etc.

d. The message may now be solved quite readily, by the usual process of converting the cipher-text letters into their plain-component equivalents and then completing the plain component sequences. The solution is as follows:

(Key: STRIKE WHILE THE IRON IS)

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

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

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

 O N I S
B O A R D N O C O M M U N I C A T I O N A F T E R T E N
A O R S L A U P A D D E R X P N B X A R I G H F X J X I

e. (1) In the foregoing illustration the steps are particularly simple because of the following circumstances:

- (a) The actual word lengths are shown.
- (b) The words are enciphered monoalphabetically by different alphabets belonging to a set of secondary alphabets.
- (c) Repetitions of plain-text words, enciphered by different alphabets, produce isomorphs and the lengths of the isomorphs are definitely known as a result of circumstance (a).

(2) Of these facts, the last is of most interest in the present connection. But what if the actual word lengths are not shown; that is, what if the text to be solved is intercepted in the usual 5-letter-group form?

SECTION IV

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS
TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, III.

	Paragraph
General remarks	14
Word separators	15
Variations and concluding remarks on foregoing systems	16

14. General remarks. - a. The cases described thus far are particularly easy to solve because the cryptanalyst has before him the messages in their true or original word lengths. But in military cryptography this is seldom or never the case. The problem is therefore made somewhat more difficult by reason of the fact that there is nothing to indicate definitely the limits of encipherment by successive keyletters. However, the solution merely necessitates more experimentation in this case than in the preceding. The cryptanalyst must take careful note of repetitions which may serve to "block out" or delimit words, and hope that when this is done he will be able to find and identify certain sequences having familiar idiomorphic features or patterns, such as those noted above. If there is plenty of text repetitions will be sufficient in number to permit of employing this entering wedge.

b. Of course, if any sort of stereotypic phraseology is employed, especially at the beginnings or endings of the messages, the matter of assuming values for sequences of cipher letters is easy, and affords a quick solution. For example, suppose that as a result of previous work it has been

the data is to handle the text digraphically, taking the 1st and 2d letters, the 2d and 3d, the 3d and 4th and so on, and distributing the final letters of each such digraph in a quadricular table. The distribution merely takes the form of tally marks, the fifth being a diagonal stroke so as to totalize the counts automatically. This yields Table 1.

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

Table 1.

c. Table 1 now presents 26 uniliteral, monoalphabetic frequency distributions. If there were more data in each distribution they could be solved in the usual manner. Naturally, the distributions which have the greatest numbers of occurrences would be the ones most easily analyzable and therefore the first to be attacked. In this process advantage would be taken of such long repetitions as are present in the cryptograms, since it is clear that given sufficient text there will be such repetitions. For example, note the 9-letter sequence underscored in message 1 which is repeated in message 3.

d. It is to be noted in connection with the subject of repetitions in this system that a repetition in the cipher text represents an equal-length repetition in the plain text. In the case of plain-text auto-keying this is not the case, the cipher-text repetition being one letter shorter in length than the plain-text repetition. Furthermore, it is to be noted that in cipher-text auto-keying, if basic enciphering equations: $\theta_{i/1} = \theta_{k/2}$; $\theta_{p/1} = \theta_{c/2}$ are in effect with a plain normal component ¹, then every time the letter A_p occurs in the plain text there will be a doublet in the cipher text, no matter what the keyletter for enciphering A_p is and no matter what the cipher component is. For example, using a direct standard alphabet as the plain component and the HYDRAULIC...XZ sequence as the cipher component (introductory keyletter X):

Key.....	X E E W A A W N N
Plain....	M A N H A T T A N
Cipher...	<u>E</u> <u>E</u> <u>W</u> <u>A</u> <u>A</u> <u>W</u> <u>N</u> <u>N</u> R

This means that doubled letters in such a system may be studied with a view to identifying certain frequently-used words containing the letter A, such as BATTALION, ARTILLERY, etc.

e. If the underscored repetition in the case of the illustrative messages is studied with this in mind, it may be assumed to represent the

¹See Par. 9b.

word ARTILLERY. Thus:

Key..... V V J L K Z E K U R . . .
 Plain.... A R T I L L E R Y . . .
 Cipher... V V J L K Z E K U R . . .

Now note that if this tentative identification is correct there are two tell-tale characteristics in the cipher equivalent. The first is the one already pointed out, namely, that the 1st letter of the cipher equivalent is the same as the cipher letter immediately preceding it. The second characteristic is the appearance of the two K's at an interval of 3 spaces. (This is, of course, a function of the particular primary components used in this particular encipherment; but it is desired merely to indicate a general principle at this point.) Hence, the word ARTILLERY is idiomorphic in this case.

Scanning the messages carefully, there will be encountered the following additional sequences which can be assumed to be the cipher equivalents of ARTILLERY because they are isomorphic with the VVJLKZEKUR sequence. Thus:

		A R T I L L E R Y
		<u>D D Q G S U N S F B</u>
In Message 1...	{	<u>V V J L K Z E K U R</u>
		<u>K K U X L O D L W T</u>
In Message 2...		<u>I I X O Z E T Z N K</u>
In Message 3...		<u>V V J L K Z E K U R</u>

If this is correct, then partial chains of equivalents in the primary cipher component may be constructed, by applying principles of indirect symmetry.

In this case it actually happens that practically the entire primary cipher component may be reconstructed from the relatively few data given by the superimposition of these hypothetical equivalents of the single word ARTILLERY. Let the student confirm the fact that the following sequence may be derived from the data given:²

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
 T E Z K R . I V F . . . Q . W G . N U S B X J D O L

²The derivation is, of course, independent of the identity of the word which the isomorphs represent.

The only missing letters are A, C, H, M, P, and Y. By testing the partial sequence to see if it is a decimation of a keyword-mixed sequence, it is found that it is derived from the keyword HYDRAULIC. Thus, while analysis of the relatively scanty individual frequency distributions contained in Table 1 would be extremely difficult, if not impossible, because of the paucity of data in the distributions, by applying the valuable tool of indirect symmetry, solution may be achieved in a relatively short time.

f. If there is a large amount of traffic in the same set of cipher alphabets, so that there is no difficulty in selecting the cipher equivalent of a single letter, say E_p , in each distribution, solution comes even more quickly than in the foregoing example.

g. The example solved in the foregoing subparagraphs offers an important lesson to the student, insofar as it teaches him that he should not immediately feel discouraged when confronted with a problem yielding only a small quantity of text and therefore affording what seems at first glance to be an insufficient quantity of data for solution. For in this example, while it is true that there are insufficient data for analysis by simple principles of frequency, it turned out that solution was achieved without any recourse to the principles of frequency of occurrence. Here, then, is one of those interesting cases of substitution ciphers of rather complex construction which are solvable without any study whatsoever of frequency distributions. Indeed, it will be found to be true that in more than a few instances the solution of quite complicated cipher systems may be accomplished not by the application of the principles of frequency, but by recourse to inductive and deductive reasoning based upon other considerations, even though the latter may often appear to be very tenuous and to rest upon quite flimsy supports.

h. Before leaving the subject of a general solution for cipher-text auto-keyed cryptograms, a final note may be added with regard to certain basic cryptographic phenomena in such cryptograms. In the illustrative example in subparagraph b, the auto-keying sequence was offset but one letter to the right of the cipher text because the introductory key consists of but a single letter. Obviously, if the initial key consists of 2, 3, 4, ... letters, the subsequent auto-keying sequence will be offset 2, 3, 4, ... letters to the right, respectively. This means that in studying cryptograms of this nature the cryptanalyst can make trials of frequency distributions of the type shown in Table I (page 51) at various intervals, in order to ascertain the length of the introductory key. Only when the proper interval is tried will the distribution be monoalphabetic. For example, in the illustrative case in subparagraphs b and c, the distribution is based upon the cipher letter immediately following each appearance of the coordinate letter.³ This interval (1) is correct in that case because the introductory key consists of one letter. But suppose the introductory key consisted of 2 letters. Then a distribution based upon the letter immediately following each appearance of the coordinate letter would no longer show monoalphabeticity, since the interval selected is not correct. In this case the distribution would have to be based upon the second letter after each appearance of the coordinate letter, and then the distribution would show monoalphabeticity. Likewise, if the introductory key consisted of 3 letters, the distribution would have to be based upon the 3d letter after each appearance of the coordinate letter, and so on.

i. Another point is of interest in connection with the external characteristics of repetitions in cryptograms of this nature. In subparagraph

³By coordinate letter is meant the letter to the left of the square which is actually the key letter.

d certain characteristics were pointed out as regards the presence of double letters in the cipher text. It is to be noted that in that case the introductory key consists of but one letter, so that whenever the letter A_p occurs in the plain text there is a double letter in the cipher text. But suppose the introductory key consisted of 2 letters. Now note what happens, using the same example as in d, but with introductory key XY.

Key.....	XY	E	W	C	W	Y	N
Plain....	M	A	N	A	T	T	A
Cipher...	E	Y	<u>N</u>	<u>C</u>	<u>W</u>	<u>Y</u>	<u>N</u>

It will be seen that the letter A_p now produces a repetition of the pattern ABA, where A and B may be any letters. (Note WCW and YNY in the example.)

If the introductory key consisted of 3 letters, then the presence of A_p would produce a repetition of the pattern ABCA, and so on.

j. Finally, the number of times patterns of the types indicated occur in specific cryptograms (AA, when the introductory key is one letter; ABA, when it is 2 letters; ABCA, when it is 3 letters, etc.), may be used as a basis for ascertaining the "base letter" in encipherment. By this is meant that, if for example, the primary components are set so that θ_k is opposite A_p , (base letter is A) and if the introductory key is say 2 letters, then the number of times the pattern ABA occurs in the cryptogram corresponds with the frequency of A_p . If the cryptogram contains 1000 letters, there should then be about 72 occurrences of the pattern ABA, since in 1000 letters of plain text there should be about 72 A's. If the base letter is T_p , then there should be about 90 occurrences of the pattern ABA, and so on. When the base letter has been ascertained by such a procedure, it is immediately possible to insert the corresponding plain-text letter throughout the text of the message. The distribution of this letter may not only serve as a check (if no inconsistencies develop) but also may lead to the assumption of values for other cipher letters.

26. Special case of solution of cipher-text auto-keyed cryptograms. -

a. Two messages with identical plain texts but different initial keywords (but of the same lengths⁴), can be solved very rapidly by reconstructing the primary components. The cryptographic texts of such messages will be isomorphic after the initial keyword portions. Note the two following superimposed messages, in which only a few identities have been indicated:

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

Starting with any pair of superimposed letters (beginning with the ninth), chains of equivalents are constructed:

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

By interpolation, these partial sequences may be united into the keyword sequence:

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

b. The initial keywords and the plain texts may now be ascertained quite easily by deciphering the messages, using this primary component slid against itself.

c. The foregoing solution affords a clue to the solution of cases in which the texts of two or more messages are not completely identical but are in part identical because they happen to have similar beginnings or endings, or contain nearly similar information or instructions. The progress in such cases is not so rapid as in the case of messages with wholly identical texts because much care must be exercised in blocking out the isomorphic sequences

⁴Two messages having identical plain texts and different length keywords can also be readily solved, but the procedure is more involved.

upon which the reconstruction of the primary components will be based.

d. (1) In the foregoing cases, the primary components used to encipher the illustrative messages were identical mixed sequences. If non-identical components are employed, the cryptograms present an interesting case for the application of a principle pointed out in a preceding text⁵.

(2) Suppose that the three messages of Par. 25b had been enciphered using a plain component different from the mixed component. The encipherments of the word ARTILLERY would still yield isomorphic sequences, from which, as has been noted, the reconstruction of the cipher component can be accomplished.

(3) Having reconstructed the cipher component (or an equivalent) the latter may be applied to the cipher text and a "decipherment" obtained. In this process any sequence of 26 letters may be used as the plain component and even the normal sequence A...Z may be employed for this purpose. The word decipherment in the next to the last sentence is enclosed by quotation marks because the letters thus obtained would not yield plain text, since the real or an equivalent plain component has not yet been found. Such "deciphered" text may be termed spurious plain text. But the important thing to note is that this text is now monoalphabetic and may be solved by the simple procedure usually employed in solving a monoalphabetic cipher produced by a single mixed alphabet. Thus, a polyalphabetic cipher may be converted to monoalphabetic terms and the problem much simplified. In other words, here is another example of the situations in which the principle of conversion into monoalphabetic terms may be applied with gratifying success. It is also an example of the dictum that the use of two differently-mixed primary components does not really give much more security than does a mixed component sliding against itself or against the normal sequence.

⁵ Military Cryptanalysis, Part I, Par. 45.

SECTION VII.

SOLUTION OF AUTO-KEY SYSTEMS, II.

	Paragraph
Preliminary remarks on plain-text auto-keying.....	27
Solution of plain-text auto-keyed cryptograms when the introductory key is a single letter.....	28
Mechanical solutions.....	29

27. Preliminary remarks on plain-text auto-keying. - a. If the cipher alphabets are unknown sequences, plain-text auto-keying gives rise to cryptograms of more intricate character than does cipher-text auto-keying, as has already been stated. As a cryptographic principle it is very commonly encountered as a new and remarkable "invention" of tyros in the cryptographic art. It apparently gives rise to the type of reasoning to which attention has been directed once before and which was then shown to be a popular delusion of the uninitiated. The novice to whom the auto-key principle comes as a brilliant flash of the imagination sees only the apparent impossibility of penetrating a secret which enfolds another secret. His reasoning runs about as follows: "In order to read the cryptogram, the would-be solver must, of course, first know the key; but the key does not become known to the would-be solver until he has read the cryptogram and has thus found the plain text. Since this is reasoning around a circle, the system is indecipherable." How unwarranted such reasoning really is in this case, and how readily the problem is solved, will be demonstrated in the next few paragraphs.

b. The discussion will, as usual, be divided into two principal cases: (1) when the cipher alphabets are known and (2) when they are unknown. Under each case there may be an introductory key consisting of a single letter, a word, or a short phrase. The single-letter initial key will be treated first.

28. Solution of plain-text auto-keyed cryptograms when the introductory key is a single letter. - a. Note the following plain-text auto-keyed encipherment of such commonly encountered plain-text words as COMMANDING, BATTALION, and DIVISION, using direct standard alphabets.

(1) B A T T A L I O N B T M T L T W B	(2) D I V I S I O N L D D A A W B
(3) C O M M A N D I N G Q A Y M N Q L V T	(4) C A P T A I N C P I T I V

These characteristics may be noted:

(1) The cipher equivalent of A_p is the plain-text letter which immediately precedes A_p . (See the two A's in BATTALION, for 1 above).

(2) A plain-text sequence of the general formula ABA yields a doublet as the cipher equivalent of the final two letters. (See IVI or ISI in DIVISION, example 2 above.)

(3) Every plain-text trigraph having A_p as its central letter yields a cipher equivalent the last two letters of which are identical with the initial and final letters of the plain-text trigraph. (See MAN in COMMANDING, example 3 above.)

(4) Every plain-text tetragraph having A_p as the initial and the final letter yields a cipher equivalent the 2d and 4th letters of which are identical with the 2d and 3d letters of the plain-text tetragraph, respectively. (See APTA in CAPTAIN, example 4 above; also ATTA in BATTALION, example 1.)

b. (1) From the foregoing characteristics and the fact that a repetition of a sequence of n plain-text letters will yield a repetition of a sequence of n-1 cipher letters, it is obvious that the simplest method of solving this type of cipher is that of the probable-word. Indeed, if the system were used for regular traffic it would not be long before the solution

would consist merely in referring to lists of cipher equivalents of commonly used words (as found from previous messages) and searching through the messages for these cipher equivalents.

(2) Note how easily the following message can be solved:

B E C J I B T M T L T W B P Q A Y M N Q H V N E T W A A L C . . .

Seeing the sequence BTMTLTWB, which is on the list of equivalents in a above (see example 1), the word BATTALION is inserted in proper position. Thus:

B E C J I B T M T L T W B P Q . . .
 B A T T A L I O N

With this as a start, the decipherment may proceed forward or backward with ease. Thus:

B E C J I B T M T L T W B P Q A Y M N Q H V N E T W A A L C . . .
 E A C H B A T T A L I O N C O M M A N D E R W I L L P L A C . . .

c. The foregoing example is based upon the so-called Vigenère method of encipherment ($\theta_{k/c} = \theta_{i/p}; \theta_{p/p} = \theta_{c/c}$). If in encipherment the plaintext letter is sought in the cipher component, its equivalent taken in the plain component ($\theta_{k/c} = \theta_{i/p}; \theta_{p/c} = \theta_{c/p}$), the steps in solution are identical, except that the list of cipher equivalents of probable words must be modified accordingly. For instance, BATTALION will now be enciphered by the sequence ZTAHLXGZ.

d. If reversed standard cipher alphabets are used, the word BATTALION will be enciphered by the sequence BHATPDUB, which also presents idiomorphic characteristics leading to the easy recognition of the word.

e. All the foregoing phenomena are based upon standard alphabets, but when mixed cipher components are used and these have been reconstructed, similar observations may be recorded and the results employed in the solution of additional messages enciphered by the same components. How messages

enciphered by unknown mixed components may be solved will be discussed presently. At this point it will merely be noted that repeated plain-text sequences of n letters produce repetitions of n-1 letters in the cipher text. Frequently-repeated trigraphs and tetragraphs, such as ENT, ING, ION, SION, TION, will produce repetitions of digraphs and trigraphs in the cipher text, so that, in addition to such repetitions of long words as are commonly encountered in military text, there will be plenty of shorter repetitions which will yield useful data. Inasmuch as the principles underlying the steps to be followed after the repetitions have been found are quite similar to those set forth under Pars. 25 and 26, in connection with cipher-text auto-keying, no example will be necessary at this point.

29. Mechanical solutions. - a. The method of solution pointed out in the preceding paragraph is based upon the successful application of the probable-word method. But sometimes the latter method fails because the commonly-expected words may not be present after all. Hence, other principles and methods, preferably mechanical in character and not dependent upon assuming the presence of any specific words may be desirable. One such method will now be described. The matter will be presented under separate headings, dependent upon the types of primary components employed.

b. The types of primary components may be classified as follows:

- (1) Primary components are identical.
 - (a) Both components progress in the same direction.
 - (b) The two components progress in opposite directions.
- (2) Primary components are different.

c. (1) Taking up the case wherein the two identical, primary components progress in the same direction, assume the following additional factors to be known by the cryptanalyst:

- (a) The primary components are both normal sequences.
- (b) The encipherment is by plain-text auto-keying.
- (c) The enciphering equations are: $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$.

(2) A message beginning QVGLB TPJTF ... is intercepted; the only unknown factor is the initial keyletter. Of course, one could try to decipher the message using each keyletter in turn, beginning with A and continuing until the correct keyletter is tried, whereupon plain text will be obtained. But it seems logical to think that all the 26 possible "decipherments" might be derived from the first one, so that the process might be much simplified, and this is true, as will now be shown. Taking the two cipher groups under consideration, let them be "deciphered" with initial keyletter A:

Cipher	<u>QVGLBTPJTF</u>
Deciphered with keyletter A	<u>QFBKRCNWXI</u>

The deciphered text is certainly not "plain text." But if one completes the sequences initiated by these letters, using the direct standard sequence for the even columns, the reversed standard for the odd columns, the plain text sequence is seen to reappear on one generatrix: It is HOSTILE FOR(CE). From this it appears that instead of going through the labor of making 26 successive trials, which would consume considerable time, all that is necessary is to have a set of strips bearing the normal direct sequence and another set bearing the reversed normal sequence, and to align the strips, alternately direct and reversed, to the first "decipherment." The plain text will now reappear on one generatrix of the completion diagram.

Initial
key
letter Q V G L B T P J T F

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

Figure 8.

(3) The peculiar nature of the phenomenon just observed, viz, a completion diagram with the vertical sequences in adjacent columns progressing in opposite directions, those in alternate columns in the same direction, calls for an explanation. Although the matter seems rather mysterious, it will not be hard to understand. First, it is not hard to see why the letters in column 1 of Fig. 8 should form the descending sequence QPQ... for these letters are merely the ones resulting from the successive "decipherment" of Q_c by the successive keyletters A,B,C,... . Now since the "decipherment" obtained from the 1st cipher letter in any row in Fig. 8 becomes the keyletter for "deciphering" the 2nd cipher letter in the same row, it is apparent that as the letters in the 1st column progress in a reversed normal (descending)

primary component (which is assumed to have been reconstructed from previous work):

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

Message: U S I N L Y Q E O P ... etc.

(2) First, the message is "deciphered" with the initial key-letter A, and then a completion diagram is established, using sliding strips bearing the mixed primary component, alternate strips bearing the reversed sequence.

Note Fig. 9, in which the plain text, HOSTILE FOR(CE), reappears on a single generatrix. Note also that whereas in Fig. 8 the odd columns contain the primary sequence in the reversed order, and the even columns contain the sequence in the direct order, in Fig. 9 the situation is reversed; the odd columns contain the primary sequence in the direct order, and the even columns contain the sequence in the reversed order. This point is brought to notice to show that it is immaterial whether the direct order is used for odd columns or for even columns; the alter-
nation in direction is all that is required in this type of solution.

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

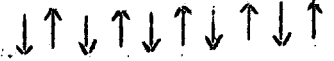


Figure 9.

e. (1) There is next to be considered the case in which the two primary components progress in opposite directions [Par. 29b(1)(b)]. Here is a message, known to have been enciphered by reversed standard alphabets, plain-text auto-keying having been followed:

X T W Z L X H Z R X

- 68 -

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

Figure 10.

paragraphs (1) and (2) hereof. Note the following solution:

Message V D D N C T S E P A ...

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

Note here that the primary mixed sequence is used for the completion sequence and that the plain text, HOSTILE FOR(CE), comes out on one generatrix. It is immaterial whether the direct or reversed mixed component is used for the completion sequence, so long as all the sequences in the diagram progress in the same direction. (See Fig. 11.)

(2) The procedure in this case is exactly the same as before, except that it is not necessary to have any alternation in direction of the completion sequences, which may be either that of the plain component or the cipher component. Note the solution in Fig. 10. Let the student ascertain why the alternation in direction of the completion sequences is not necessary in this case.

(3) In the foregoing case the alphabets were reversed standard, produced by the sliding of the normal sequence against its reverse. But the underlying principle of solution is the same even if a mixed sequence were used instead of the normal; so long as the sequence is known, the procedure to be followed is exactly the same as demonstrated in sub-

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

Figure 11.

f. (1) There remains now to be considered only the case in which the two components are different mixed sequences. Let the two primary components be 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	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X

and the message:

C F U Y L V X U D J

(2) First "decipher" the message with any arbitrarily selected initial keyletter, say A, and complete the plain component sequence in the first column. (Fig. 12a)

Cipher...	C F U Y L V X U D J
Plain....	<u>L F Q X W X A W S F</u>
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	

Figure 12a.

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

Figure 12b

Now prepare a strip bearing the cipher component reversed, and set it below the plain component so that $F_p = L_c$, a setting given by the 1st two letters of the spurious "plain text" recovered. 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...	F X O N W L A V K M U J D T H E S G I Z Q C R Y P B

(3) Now opposite each letter of the completion sequence in column 1, write its plain-component equivalent, as given by the juxtaposed sequences above. This gives what is shown in Fig. 12b. Then reset the two sequences (reversed cipher component and the plain component) so that $Q_p = F_c$ (to correspond with the 2d and 3d letters of the spurious plain text); write down the plain-component equivalents of the letters in column 2, forming column 3. Continue this process, scanning the generatrices from time to time, resetting the two components and finding equivalents from column to

- 71 -

column, until it becomes evident on what generatrix the plain text is reappearing. In Fig. 12c it is seen that the plain text generatrix is the one beginning HOST, and from this point on the solution may be obtained directly, by using the two primary components.

(4) When the plain component is also a mixed sequence (and different from the cipher component), the procedure is identical with that outlined in subparagraphs (1)-(3) above. The fact that the plain component in the preceding case is the normal sequence is of no particular significance in the solution, for it acts as a mixed sequence would act under similar circumstances. To demonstrate, suppose the two following components were used in encipherment of the message below:

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

To solve the message, "decipher" the text with any arbitrarily selected initial keyletter and proceed exactly as in subparagraphs (2) and (3) above.

Thus: Cipher B B V Z U D Q X J D
 "Plain" ($\theta_k=X$) ... V Y R I Y Z E F O R

Note the completion diagram in Fig. 13 (page 72), which shows the word HOST... very soon in the process. From this point on the solution may be obtained directly, by using the two primary components.

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

Figure 12c.

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

g. (1) Another "mechanical" solution for the foregoing cases will now be described because it presents rather interesting cryptanalytic sidelights. Take the message R E F E R E N C E H I S P R E F E R E N C E I N R E F E R E N C E B O O K S A N D R E F E R E N C E C H A R T S . . . and encipher it by plain-text auto-key, with normal direct primary components, initial key setting $A_p = G_c$. Then note the underscored repetitions:
 R E F E R E N C E H I S P R E F E R E N C E I N
 X V J J V V R P G L P A H G V J J V V R P G M V
 R E F E R E N C E B O O K S A N D R E F E R E N C E
 E V J J V V R P G F P C Y C S N Q U V J J V V R P G
 C H A R T S
 G J H Y K L

Figure 13.

Now suppose the message has been intercepted and is to be solved. The only unknown factor will be assumed to be the initial keyletter. Let the message be "deciphered" by means of any initial keyletter, say A, and then note the underscored repetitions in the spurious plain text.

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

The original four 8-letter repetitions now turn out to be two different sets of 9-letter repetitions, which seems rather peculiar. But now let the spurious plain text with its real plain text, and the cipher text, be transcribed as though one were dealing with a periodic cipher involving two alphabets, as shown herewith. It will here be seen that the letters in column 1 are mono-alphabetic, and so are those in column 2. In other words, an auto-key cipher,

which is commonly regarded as a polyalpha-	1-2	1-2	1-2	1-2
betic, aperiodic cipher, has been converted	R E	E F	R E	E F
into a 2-alphabet, periodic cipher, the indi-	X Y	K Z	X Y	K Z
vidual alphabets of which are now monoalpha-	F E	E R	N C	E R
betic in nature. The two repetitions of	L Y	K L	T W	K L
X Y L Y X Y T W K represent encipherments	R E	E N	E B	E N
of the word REFERENCE, in alphabets 1-2-	X Y	K H	K V	K H
1-2-1-2-1-2-1; the two repetitions of	N C	C E	O O	C E
L K Z K L K H I Y likewise represent en-	T W	I Y	U I	I Y
cipherments of the same word but in alpha-	E H	I N	K S	C H
bets 2-1-2-1-2-1-2-1-2.	K B	O H	Q M	I B

(2) Later on it will be seen how this method of converting an auto-key cipher into a periodic cipher may be applied to the case where an introductory keyword is used as the initial keying element instead of a single letter, as in the present case. At this point let the student try to figure out for himself the basis for the phenomena demonstrated above. A hint will be found by referring to the case described under c under this paragraph.

h. (1) In the foregoing case the primary components were identical normal sequences progressing in the same direction. If they were mixed sequences the phenomena observed above would still hold true, and so long as the sequences are known, the indicated method of solution may be applied.

(2) When the two primary components are known but differently mixed sequences, this method of solution is too involved to be practical. It is more practicable to try successive initial keyletters, noting the plain text each time and resetting the strips until the correct setting has been ascertained, as will be evidenced by obtaining intelligible plain text.

SECTION VIII.

SOLUTION OF AUTO-KEY SYSTEMS, III.

	Paragraph
Solution of plain-text auto-keyed cryptograms when the introductory key is a word or phrase	30
Subsequent steps after determining the length of the introductory key	31
Conversion of foregoing aperiodic cipher into periodic form ...	32
Concluding remarks on auto-key systems	33

30. Solution of plain-text auto-keyed cryptograms when the introductory key is a word or phrase. - a. In the foregoing discussion of plain-text auto-keying, the introductory key was assumed to consist of a single letter, so that the subsequent keyletters are displaced one letter to the right with respect to the text of the message itself. But sometimes a word or phrase may serve this function, in which case the subsequent key is displaced as many letters to the right of the initial plain-text letter of the message as there are letters in the initial key. This will not, as a rule, interfere in any way with the application of the principles of solution set forth in Par. 28 to that part of the cryptogram subsequent to the introductory key, and a solution by the probable-word method and the study of repetitions can be reached. However, it may happen that trial of this method is not successful in certain cryptograms because of the paucity of repetitions, or because of failure to find a probable word in the text. When the cipher alphabets are known there is another point of attack which is useful and interesting. The method consists in finding the length of the introductory key and then solving by frequency principles. Just how this is accomplished will now be explained.

b. Suppose that the introductory keyword is HORSECHESNUT, that the plain-text message is as below, and that identical primary components progressing in the same direction are used to encipher the message, by encipher-equation $\theta_{k/c} = \theta_{i/p}; \theta_{p/p} = \theta_{c/c}$. Let the components be the normal sequence.

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

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

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

Key...	T	F	R	O	N	T	A	N	D	C	O	N	C	E	N	T	R	A	T	I	N	G	A	R	T	I
Plain..	E	N	T	R	A	T	I	N	G	A	R	T	I	L	L	E	R	Y	T	H	E	R	E	X	W	I
Cipher.	X	S	K	F	N	M	I	A	J	C	F	G	K	P	Y	X	I	Y	M	P	R	X	E	O	P	Q

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

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

It will now be noted that since the introductory key contains 13 letters the 14th letter of the message is enciphered by the 1st letter of the plain text, the 15th by the 2d, and so on. Likewise, the 27th letter is enciphered by the 14th, the 28th by the 15th, and so on. Hence, if the 1st cipher letter is deciphered, this will give the key for deciphering the 14th, the latter will give the key for the 27th, and so on. An important step in the solution of a message of this kind would therefore involve ascertaining the length of the introductory key. This step will now be explained.

c. Since the plain text itself constitutes the keyletters in this system (after the introductory key), these keyletters will occur with their normal frequencies, and this means that there will be many occurrences of E, T, O, A, N, I, R, S enciphered by E_k ; there will be many occurrences of these same high-frequency letters enciphered by T_k , by O_k , by A_k , and so on. In fact, the number of times each of these combinations will occur may be calculated statistically. With the enciphering conditions set forth

under \underline{b} above¹ E_p enciphered by T_k , for example, will yield the same cipher equivalent as T_p enciphered by E_k ; in other words two encipherments of any pair of letters of which either may serve as the key for enciphering the other must yield the same cipher resultant. It is the cryptographic effect of these two phenomena working together which permits of ascertaining the length of the introductory key in such a case. For every time a given letter, θ_p , occurs in the plain text it will occur \underline{n} letters later as a keyletter, θ_k , and \underline{n} in this case equals the length of the introductory key. Note the following illustration:

	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) Key.....	H	O	R	S	E	C	H	E	S	T	N	U	T	T
(2) Plain....	T	E
(3) Cipher...	X

	1	2	3	4	5	6	7	8	9
	E
	T
	X

Here it will be noted that E_p in line (2) has a T_p on either side of it, at a distance of 13 intervals; the first encipherment (E_p by T_k) yields the same equivalent (X_c) as the second encipherment (T_p by E_k). Two cipher letters are here identical, at an interval equal to the length of the introductory key. But the converse is not true; that is, not every pair of identical letters in the cipher text represents a case of this type. For in this system identity in two cipher letters may be the result of the following three conditions each having a statistically ascertainable probability of occurrence:

- (1) A given plain-text letter is enciphered by the same key letter two different times, at an interval which is purely accidental; the cipher equivalents are identical but could not be used to give any information about the length of the introductory key.

¹It is important to note that the two components must be identical sequences and progress in the same direction. If this is not the case, the entire reasoning is inapplicable.

- (2) Two different plain-text letters are enciphered by two different keyletters; the cipher equivalents are fortuitously identical.
- (3) A given plain-text letter is enciphered by a given key letter and later on the same plain-text letter serves to encipher another plain-text letter which is identical with the first key letter; the cipher equivalents are causally identical.

It can be proved that the probability for identities of the third type is greater than that for identities of either or both 1st and 2d types for that interval which corresponds with the length of the introductory key; that is, if a tabulation is made of the intervals between identical letters in such a system as the one being studied, the interval which occurs most frequently should coincide with the length of the introductory key. The demonstration of the mathematical basis for this fact is beyond the scope of the present text; but a practical demonstration will be convincing.

d. Let the illustrative message be transcribed in lines of say 11, 12, and 13 letters, as in Fig. 14.

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

abc

Figure 14.

- 78 -

In each transcription, every pair of superimposed letters is noted and the number of identities is indicated by ringing the letters involved, as shown above. The number of identities for an assumed introductory-key length 13 is 9, as against 3 for the assumption of a key of 11 letters, and 5 for the assumption of a key of 12 letters.

e. Once having found the length of the introductory key, two lines of attack are possible: the composition of the key may be studied, which will yield sufficient plain text to get a start toward solution; or, the message may be reduced to periodic terms and solved as a repeating-key cipher. The first line of attack will be discussed first, it being constantly borne in mind in this paragraph that the entire discussion is based upon the assumption that the cipher alphabets are known alphabets. The illustrative message of b above will be used.

31. Subsequent steps after determining the length of the introductory key. - a. Assume that the first letter of the introductory key is A and decipher the 1st cipher letter T_0 (with direct standard alphabets). This yields T_p and the latter becomes the keyletter for the 14th letter of the message. The 14th letter is deciphered: $D_c(T_k) = K_p$; the latter becomes the keyletter for the 27th letter and so on, down the entire first column of the message as transcribed in lines of 15 letters. The same procedure is followed using B as the initial keyletter, then C, and so on. The message as it appears for the first three trials (assuming A, B, then C as the initial keyletter) is shown in Fig. 15. (page 79).

1 2 3 4 5 6 7 8 9 10 11 12 13
 T M C W J V M P S G X C L
 T
 D
 K
 P
 F
 C
 X
 R
 U
 G
 M
 X
 L
 P
 E
 W
 S
 C
 K
 V
 L
 E
 T

(a)

1st Column of Fig. 14(c)
 "deciphered" with initial
 $\theta_k = A.$

1 2 3 4 5 6 7 8 9 10 11 12 13
 T M C W J V M P S G X C L
 S
 D
 L
 P
 E
 C
 Y
 R
 T
 G
 N
 X
 K
 P
 F
 W
 R
 C
 L
 V
 K
 E
 U

(b)

1st Column of Fig. 14(c)
 "deciphered" with initial
 $\theta_k = B.$

Note: In each column of the diagram the odd letters are the cipher letters in the corresponding column of Fig. 14c; the even letters in the column are decipherments of the odd letters, by the keyletter shown at the bottom of the diagram. (See Par. 31a.)

1 2 3 4 5 6 7 8 9 10 11 12 13
 T M C W J V M P S G X C L
 R
 D
 M
 P
 D
 C
 Z
 R
 S
 G
 O
 X
 J
 P
 G
 W
 Q
 C
 M
 V
 J
 E
 V

(c)

1st Column of Fig. 14(c) "deciphered"
 with $\theta_k = C.$

Figure 15.

b Inspection of the results of these three trials soon shows that the entire series of 26 trials need not be made, for the results can be obtained from the very first trial. This may be shown graphically by superimposing merely the results of the first three trials horizontally. Thus:

Cipher Letters of Col. 1 - Fig. 15 -	T D P C R G X P W C V E
A -	T K F X U M L E S K L T
B -	S L E Y T N K F R L K U
Key Letters C -	R M D Z S O J G Q M J V
D -	↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓
E -	↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓

Figure 16.

c. It will be noted that the vertical sequences in adjacent columns proceed in opposite directions, whereas those in alternate columns proceed in the same direction. The explanation for this alternation in progression is the same as in the previous case wherein this phenomenon was encountered

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

(Par. 29c(2)), and the sequences in Fig. 16 may now be completed very quickly. The diagram becomes as shown in Fig. 17.

d. One of the horizontal lines or generatrices of Fig. 17 is the correct one; that is, it contains the actual plain-text equivalents of the 1st, 14th, 27th, ... letters of the message. The correct generatrix can be selected by mere ocular examination, as is here possible (see generatrix marked by asterisk in Fig. 17), or it may be selected by a frequency test, assigning weights to each letter according to its normal plain-text frequency. (See Par. 14f of Military Cryptanalysis, Part II.)

Figure 17.

- 81 -

e. Identical procedure is followed with respect to column 2, 3, 4, ... of Fig. 14c, with the result that the initial keyword HORSECHESTNUT is reconstructed and the whole message may be now deciphered quite readily.

32. Conversion of foregoing aperiodic cipher into periodic form. - a. In Par. 30e it was stated that an aperiodic cipher of the foregoing type may be reduced to periodic terms and solved as though it were a repeating-key cipher, provided the primary components are known sequences. The basis of the method lies in the phenomena noted in Par. 29. An example will be given.

b. Let the cipher text of the message of Par. 30b be set down again, as in Fig. 14c:

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

Using direct standard alphabets (Vigenère method), "decipher" the second line by means of the first line, that is, taking the letters of the second line as cipher text, those of the first line as keyletters. Then use the thus-found "plain text" as "key letters" and "decipher" the third line of Fig 14c, as shown in Fig. 18. Thus:

"Key".....	T	M	C	W	J	V	M	P	S	G	X	C	L
Cipher....	D	C	N	I	N	O	N	Y	G	U	O	I	N
"Plain"...	K	Q	L	M	E	T	Z	J	O	O	R	G	C
"Key".....	K	Q	L	M	E	T	Z	J	O	O	R	G	C
Cipher....	P	E	T	X	Q	G	T	R	X	F	J	I	M
"Plain"...	F	O	I	L	M	N	U	I	J	R	S	C	K

Figure 18.

- 82 -

Continue this operation for all the remaining lines of Fig. 14c and write down the results in lines of 26 letters. Thus:

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

Figure 19.

Now write down the real plain text of the message in lines of 26 letters. Thus:

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

Figure 20.

c. When the underlined repetitions in Figs. 19 and 20 are compared, they are found to be identical in the respective columns, and if the columns of Fig. 19 are tested, they will be found to be monoalphabetic. The cipher message now gives every indication of being a repeating-key cipher. It is not difficult to explain this phenomenon in the light of the demonstration given in Par. 29g. First, let the keyword HORSECHESTNUT be enciphered by the following alphabet:

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

"Plain".... H O R S E C H E S T N U T
 "Cipher"... T M J I W Y T W I H N G H

Then let the message MY LEFT FLANK etc., be enciphered by direct standard alphabets as before, but for the key add the monoalphabetic equivalents of HORSECHESTNUT (TMJIW...) to the key itself, that is, use the 26-letter key HORSECHESTNUTTMJIWYTWIHNHGH in a repeating-key manner. Thus (in Fig. 21):

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

Figure 21.

The cipher resultants of this process of enciphering a message coincide exactly with those obtained from the "deciphering" operation that gave rise to Fig. 19. How does this happen?

d. First, let it be noted that the sequence TMJI ... , which forms the second half of the key for enciphering the text in Fig. 21 may be described as the standard alphabet complement of the sequence HORSECHESTNUT, which forms the first half of that key. Arithmetically, the sum of a letter of the first half and its homologous letter in the second half is 26. Thus:

$$\begin{aligned}
 H + T &= 7 + 19 = 26 = 0 \\
 O + M &= 14 + 12 = 26 = 0 \\
 R + J &= 17 + 9 = 26 = 0 \\
 S + I &= 18 + 8 = 26 = 0 \\
 E + W &= 4 + 22 = 26 = 0
 \end{aligned}$$

That is, every letter of HORSECHESTNUT plus its homologous letter of the sequence TMJIWYTWIHNGH equals 26, which is here the same as zero. In other words, the sequence TMJIWYTWIHNGH is, by cryptographic arithmetic, equivalent to "minus HORSECHESTNUT." Therefore, in Fig. 21, enciphering the second half of each line by the keyletters TMJIWYTWIHNGH (i.e., adding 19, 12, 9, 8, ...) is the same as deciphering by the keyletters HORSECHESTNUT (i.e., subtracting

- 84 -

7, 14, 17, 18, ...). For example:

$$R_p (T_k) = 17 + 19 = 36 = 10 = K, \text{ and}$$

$$R_p (-H_k) = 17 - 7 = 10 = K$$

$$E_p (M_k) = 4 + 12 = 16 = Q_c, \text{ and}$$

$$E_p (-O_k) = 4 - 14 = (26 + 4) - 14 = 16 = Q_c, \text{ and so on.}$$

e. Refer now to Fig. 19. The letters in the first half of line 1, beginning TMCWJ ... are identical with those in the first half of line 1 of Fig. 21. They must be identical because they are produced from identical elements. The letters in the second half of this same line in Fig. 19, beginning KQLME ... were produced by deciphering the letters in the second line of Fig. 14c. Thus (taking for illustrative purposes only the first five letters in each case):

$$K Q L M E = D C N I N - T M C W J$$

$$\text{But } D C N I N = R E C E I + M Y L E F$$

$$\text{And } T M C W J = M Y L E F + H O R S E$$

$$\text{Hence, } K Q L M E = (R E C E I + M Y L E F) - (M Y L E F + H O R S E)$$

$$\text{Or, } K Q L M E = R E C E I - H O R S E \quad (1)$$

As for the letters in the second half of line 1 of Fig. 21, also beginning KQLME ... , these letters were the result of enciphering RECEI by TMJIW.

Thus:

$$K Q L M E = R E C E I + T M J I W$$

But it has been shown in subpar. d above that

$$T M J I W = -H O R S E$$

$$\text{Hence, } K Q L M E = R E C E I + (-H O R S E)$$

$$\text{Or, } K Q L M E = R E C E I - H O R S E \quad (2)$$

Thus, equations (1) and (2) turn out to be identical but from what appear to be quite diverse sources.

f. What has been demonstrated in connection with the letters in line 1 of Figs. 19 and 21 holds true for the letters in the other lines of these two figures, and it is not necessary to repeat the explanation. The steps show that the originally aperiodic, auto-key cipher has been converted, through a knowledge of the primary components, into a repeating-key cipher with a period twice the length of the introductory key. The message may now be solved as an ordinary repeating-key cipher.

g. (1) The foregoing case is based upon encipherment by the enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$. When encipherment by the enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/c} = \theta_{c/p}$ has been followed, the conversion of a plain-text auto-keyed cipher yields a repeating-key cipher with a period equal to the length of the introductory key. In this conversion, the enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$ are used in finding equivalents.

(2) An example may be useful. Note the encipherment of the following message by auto-key method by enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/c} = \theta_{c/}$

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

(3) If the message is written out in lines corresponding to the length of the introductory key, and each line is enciphered by the one directly above it, using the enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$ in finding equivalents, the results are as shown in Fig. 22b. But if the same message is enciphered by equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/c} = \theta_{c/p}$, using the word TUESDAY as a repeating key, the cipher text (Fig. 22c) is identical with that obtained in Fig. 22b by enciphering each successive line with the line above it.

- 86 -

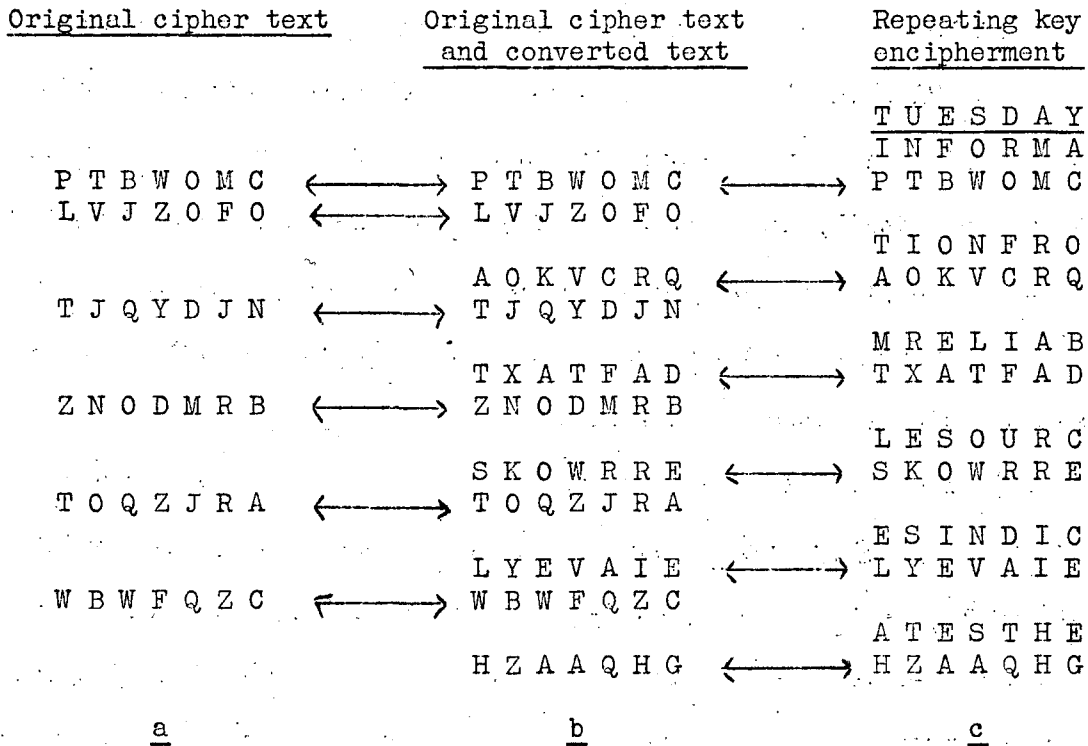


Figure 22.

(4) Now note that the sequences joined by arrows in Fig. 22 b and c are identical and since it is certain that Fig. 22c is periodic in form because it was enciphered by the repeating-key method, it follows that Fig. 22b is now also in periodic form, and in that form the message could be solved as though it were a repeating-key cipher.

h (1) In case of primary components consisting of a direct normal sequence sliding against a reversed normal (U. S. Army disk), the process of converting the auto-key text to periodic terms is accomplished by using two direct normal sequences and "deciphering" each line of the text (as transcribed in periods) by the line above it. For example, here is a message auto-enciphered by the aforementioned disk, with the initial keyword TUESDAY:

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

(2) The cipher text is transcribed in periods equal to the length of the initial keyword (7 letters) and the 2d line is "deciphered" with key-letters of the 1st line, using enciphering equations $\theta_{k/c} = \theta_{i/p}$; $\theta_{p/p} = \theta_{c/c}$. The resultant letters are then used as keyletters to "decipher" the 3d line of text and so on. The results are as seen in Fig. 23b. Now let the original message be enciphered in repeating-key manner by the disk, with the keyword TUESDAY, and the result is Fig. 23c. Note that the odd or alternate lines of Figs. 23b and c are identical, showing that the auto-key text has been converted into repeating-key text.

<u>Original cipher text</u>		<u>Original cipher text and converted text</u>		<u>Repeating-key encipherment</u>
				<u>T U E S D A Y</u>
				I N F O R M A
L H Z E M O Y	↔	L H Z E M O Y	↔	L H Z E M O Y
P F R B M V M	↔	P F R B M V M		T I O N F R O
		A M Q F Y J K	↔	A M Q F Y J K
H R K C X R N	↔	H R K C X R N		M R E L I A B
		H D A H V A X	↔	H D A H V A X
B N M X O J Z	↔	B N M X O J Z		L E S O U R C
		I Q M E J J W	↔	I Q M E J J W
H M K B R J A	↔	H M K B R J A		E S I N D I C
		P C W F A S W	↔	P C W F A S W
E Z E V K B Y	↔	E Z E V K B Y		A T E S T H E
		T B A A K T U	↔	T B A A K T U
<u>a</u>		<u>b</u>		<u>c</u>

Figure 23.

i. The foregoing procedures indicate a simple method of solving ciphers of the foregoing types, when the primary components or the secondary cipher alphabets are known. It consists in assuming introductory keys of various lengths, converting the cipher text into repeating-key form, and then examining the resulting diagrams for repetitions. When a correct key length is

assumed, repetitions will be as numerous as should be expected in ciphers of the repeating-key class; incorrect assumptions for key length will not show so many repetitions.

j. All the foregoing presupposes a knowledge of the cipher alphabets involved. When these are unknown, recourse must be had to first principles and the messages must be solved purely upon the basis of probable words, and repetitions, as outlined in Pars. 27-28.

38. Concluding remarks on auto-key systems. - a. Both cipher-text and plain-text auto-keying as cryptographic methods have, as shown, serious weaknesses which exclude them from practical usage in military cryptography. Besides being comparatively slow and subject to error, they are rather easily solvable, even when unknown cipher alphabets are employed.

b. In both systems there are characteristics which permit of identifying a cryptogram as belonging to this class of substitution. Both cases will show many repetitions in the cipher text, but in cipher-text auto-keying there will be fewer repetitions than in the original plain text, whereas in plain-text auto-keying there will be as many repetitions in the cipher text as in the original plain text. In both cases, repetitions in the cipher text are shorter than the equivalent repetitions in the plain text, the difference corresponding to the number of letters in the introductory key. These repetitions will show no constancy as regards intervals between them, and a uniliteral frequency distribution will show such messages to be polyalphabetic in nature. Cipher-text auto-keying may be distinguished from plain-text auto-keying by the appearance of the frequency distribution of the second member of sets of two letters separated by the length of the introductory key (see Par. 25h). In the case of cipher-text auto-keying these frequency distributions will be monoalphabetic in nature; in plain-text auto-keying such frequency distributions will not show monoalphabetic characteristics.

SECTION IX

METHODS OF LENGTHENING OR EXTENDING THE KEY.

	Paragraph
Preliminary remarks	34
Extended and nonrepeating keys; the so-called "running-key system"	35
Other systems employing lengthy keying sequences	36

34. Preliminary remarks. - In Par. 1b of this text it was stated that two procedures suggest themselves for eliminating the weaknesses introduced by periodicity of the type produced by simple, repeating-key methods. The first of these, when studied, embraced some of the very simple methods of suppressing or destroying periodicity, by such devices as interrupting the key and using variable length groupings of plain text. It was demonstrated that subterfuges of this simple nature are inadequate to eliminate the weaknesses referred to, and must be discarded in any system intended to afford real security. The other alternative suggested in Par. 1b therefore remains now to be investigated, viz, that of lengthening the keys to a point where there would seem to be an insufficient amount of text to enable the cryptanalyst to solve the traffic. Attempts toward this end usually consist in extending the key to such a length that the enemy cryptanalysts will have only a very limited number of periods to work with. The key may, indeed, be lengthened to a point where it becomes as long as, or longer than, the text to be enciphered, so that the key is used only once.

35. Extended and nonrepeating keys. - a. It is obvious that one of the simplest methods of lengthening the key to a message is to use a long phrase or even a complete sentence, provided it is not too long to remember. In addition to the difficulties that would be encountered in practical military cryptography in selecting long mnemonic phrases and sentences which would have to be imparted to many clerks, there is the fact that the probable-word

method of solution still remains as a powerful tool in the hands of enemy cryptanalysts. And if only a word or two of the key can be reconstructed as a result of a fortunate assumption, it is obvious that the enemy cryptanalysts could readily guess the entire key from a fragment thereof, since any long phrase or sentence which is selected because it can easily be remembered is likely to be well known to many people.

b. There are, however, more or less simple methods of employing a short mnemonic key in order to produce a much longer key. Basically, any method of transposition applied to a single alphabetic sequence repeated several times will yield a fairly long key, which, moreover, has the advantage of being unintelligible and thus approaching a random selection of letters. For example, a numerical key may be derived from a word or a short phrase; this numerical key may then be applied as a columnar-transposition key for a rectangle within which the normal alphabet has been repeated a previously agreed upon number of times in a normal (left to right) or prearranged manner. The letters when transcribed from the transposition rectangle then become the successive letters for enciphering the plain text, using any desired type of primary components. Or, if a single transposition is not thought to be sufficiently secure, a double transposition will yield a still more mixed up sequence of key letters. Other types of transposition may be employed for the purpose, including various kinds of geometric figures. Also, a non-transposition method of lengthening the keying sequence and at the same time introducing an irregularity, such as aperiodic interruption has already been described (see Par. 18).

c. Another method of developing a long key from a short mnemonic one is that shown below. Given the keyword CHRISTMAS, a numerical sequence is first derived and then one writes down successive sections of this numerical key, these sections terminating with the successive numbers 1, 2, 3, ... of

the numerical key. Thus:

Mnemonic Key ... C H R I S T M A S
 Numerical key... 2-3-6-4-7-9-5-1-8

Extended key: C H R I S T M A | C | C H | C H R I | C H R I S T M | C H R |
 1 | 2 | 3 | 4 | 5 | 6 |
 7 | 8 | 9 |
 C H R I S | C H R I S T M A S | C H R I S T |

Thus the original key of only 9 letters is expanded to one of 45 letters ($1 + 2 + 3 + \dots + 9 = 45$). The longer key is also an interrupted key of the type noted under Par. 17, but if the message is long enough to require several repetitions of the expanded key the encipherment becomes periodic and can be handled by the usual methods employed in solving repeating-key ciphers. If the basic key is fairly long, so that the expanded key becomes a quite lengthy sequence, then the message or messages may be handled in the manner explained in Par. 20.

36. Other systems employing lengthy keying sequences. - a. The so-called "running-key" system. To be mentioned in connection with this subject of extensive or lengthy keys is the cipher system known as the running-key, continuous-key, or nonrepeating-key system, in which the plain text of a previously-agreed-upon book serves as the source for successive keyletters for encipherment.¹ The solution of this type of cipher, an accomplishment which was once thought impossible, presents some interesting phases and will be considered shortly. At this point it is merely desired to indicate that according to the running-key system the key for an individual message may be as long as the message and never repeat; but if a large group of communicants employ the same book, it may happen that there will be several messages in the same key and they will all begin with the same initial keyletter; or,

¹Sect. IX, Advanced Military Cryptography.

there will be several which will "overlap" one another with respect to the key, that is, they will begin with different initial keyletters but soon will fall under the same sequence of keyletters.

b. The so-called "progressive-alphabet" system. In the so-called progressive-alphabet system the basic principle is quite simple. Two or more primary elements are arranged or provided for according to a key which may be varied from time to time; the interaction of the primary elements results in making available for cryptographic purposes a set of cipher alphabets; all the latter are employed in a fixed sequence or progression; hence the designation progressive-alphabet system. If the number of alphabets available for such use is rather small, and if the text to be enciphered is much longer than the sequence of alphabets, then the system reduces to a periodic method. But if the number of alphabets is large, so that the sequence is not repeated, then of course, the cryptographic text will exhibit no periodic phenomena.

c. The series of cipher alphabets in such a system constitutes a keying sequence. Once set up, often the only remaining element in the key for a specific message is the starting point in the sequence, that is, the initial cipher alphabet employed in enciphering a given message. If this keying sequence must be employed by a large group of communicants, and if all messages employ the same starting point in the keying sequence, obviously the cryptograms may simply be superimposed without any preliminary testing to ascertain proper points for superimposition. The student has already been shown how cases of this sort may be solved. However, if messages are enciphered with varying starting points, the matter of superimposing them properly takes on a different aspect. This will soon be treated in detail.

d. In addition to the foregoing, there are, of course, a great many mechanical methods of producing a long key, such as those employed in mechani-

- 93 -

cal or electrical cipher machines. In most cases these methods depend upon the interaction of two or more short, primary keys which jointly produce a single, much longer, secondary or resultant key. (See Par. 4). Only brief reference can be made at this point in the cryptanalytic studies to cases of this kind. A detailed treatment of complex examples would require much time and space so that it will be reserved for subsequent texts.

e. Finally, there must be mentioned certain devices in which, as in encipherment by the auto-key method, the text itself serves to produce the variation in cipher equivalents, by controlling the selection of secondary alphabets, or by influencing or determining the sequence with which they will be employed. Naturally, in such cases the key is automatically extended to a point where it coincides in length with that of the text. An excellent example of such a device is that known as the Wheatstone², the solution of which will be described in its proper place.

²See Sect. XII, Advanced Military Cryptography.

SECTION X

GENERAL PRINCIPLES UNDERLYING SOLUTION OF SYSTEMS
EMPLOYING LONG OR CONTINUOUS KEYS.

	Paragraph
Solution when the primary components are known sequences.....	37
Solution of a running-key cipher when the primary components are known	38
Solution of a progressive-alphabet cipher when the primary components are known	39
General solution for ciphers involving a long keying sequence of fixed length and composition	40

37. Solution when the primary components are known sequences. - a. As usual, the solution of cases involving long or continuous keys will be treated under two headings: first, when the primary components are known sequences; second, when these elements are wholly unknown or partially unknown.

b. Since the essential purpose in using long keys is to prevent the formation of repetitive cycles within the text, it is obvious that in the case of very long keying sequences the cryptanalyst is not going to be able to take the text and break it up into a number of small cycles which will permit the establishment of monoalphabetic frequency distributions that can readily be solved, an end which he can attain all the more readily if to begin with he knows the primary sequences. But, there nearly always remains the cryptanalyst's last resort: the probable-word method. Inasmuch as this method is applicable to most of these cases, even to that of the running-key system, which perhaps represents the furthest extension of the principle of long keying sequences, an example using a cryptogram of the latter type will be studied.

38. Solution of a running-key cipher when the primary components are known. - a. In Par. 35c mention was made of the so-called running-key, continuous-key, or nonrepeating-key system, in which the plain text of a previously-agreed-upon book serves as the source for successive keyletters for

encipherment. Since the running-key system is entirely aperiodic, and the cipher text can therefore not be arranged in superimposed short cycles, as in the case of the repeating-key system, it would appear on first consideration to be "indecipherable" without the key.¹ But if the student will bear in mind that one of the practical methods of solving a repeating-key cipher is that of the probable word², he will immediately see that the latter method can also be applied in solving the nonrepeating-key system. The essence of the matter is this: the cryptanalyst may assume the presence of a probable word in the text of the message; if he knows the primary components involved, and if the assumed word actually exists in the message, he can locate it by checking against the key, since the latter is intelligible text. Or, he may assume the presence of a probable word or even of a phrase such as "to the", "of the", etc., in the key text and check his assumption against the text of the message. Once he has forced such an entering wedge into either the message or the key, he may build upon this foundation by extending his assumptions for text alternately in the key and in the message, thus gradually reconstructing both. For example, given a cryptogram containing the sequence ... HVGGLWBESLTR ..., suppose he assumes the presence of the phrase THAT THE in the key text and finds a place in the plain text where this yields MMUNITI. Thus, using reversed standard cipher alphabets:

¹At one time, indeed, this view was current among certain cryptographers, who thought that the principle of factoring the intervals between repetitions in the case of the repeating-key cipher formed the basis for the only possible method of solving the latter type of system. Since, according to this erroneous idea, factoring cannot be applied in the case of the running-key system (using a book as the key), therefore no solution is possible. How far this idea is from the truth will presently be seen.

²See Military Cryptanalysis, Part II, Par. 25.

Assumed key textT H A T T H E ...
 Cipher text..... ..H V G G L O W B E S L T R ...
 Resultant plain textM M U N I T I ...

This suggests the word AMMUNITION. The ON in the cipher text then yields PR as the beginning of the word after THE in the key text. Thus:

Assumed key textT H A T T H E P R ...
 Cipher textH V G G L O W B E S L T R ...
 Resultant plain textM M U N I T I O N ...

PR must be followed by a vowel, with O the most likely candidate. He finds that O yields W in the plain text, which suggests the word WILL. The latter then yields OTEC in the key, making the latter read THAT THE PROTEC... . Thus:

Assumed key text..... ..T H A T T H E P R O T E C ...
 Cipher textH V G G L O W B E S L T R ...
 Resultant plain textM M U N I T I O N W I L L ...

This suggests the word PROTECTION, PROTECTIVE, PROTECTING, etc. Thus extending one text a few letters serves to "coerce" a few more letters out of the other, somewhat as in the case of two boys who are running approximately abreast in a race; as soon as one boy gets a bit ahead the spirit of competition causes the other to overtake and pass the first one; then the latter puts forth a little more effort, overtakes and passes the second boy. Thus the boys alternate in overtaking and passing each other until the race is run. The only point in which the simile fails is that while the boys usually run forward all the time, that is, in a single direction, the cryptanalyst is free to work in two directions -- forward and backward from an internal point in the message. He may, in the case of the example cited above, continue his building up process by adding A to the front of MMUNITI as well as ON to the rear. If he reaches the end of his resources on one end, there remains the other end for experimentation. He is certainly unlucky if both ends terminate in complete words both for the message and for the key, leaving him without a single clue to the next word in either, and forcing him to a more intensive use of his

imagination, guided only by the context.

b. In the foregoing illustration the cryptanalyst is assumed to have only one message available for his experimentation. But if he has two or more messages which either begin at identical initial points with reference to the key, or overlap one another with respect to the key, the reconstruction process described above is, of course, much easier and is accomplished much more quickly. For if the messages have been correctly superimposed with reference to the key text, the addition of one or two letters to the key yields suggestions for the assumption of words in several messages. The latter lead to the addition of several letters to the key, and so on, in an ever-widening circle of ideas for further assumptions, since as the process continues the context affords more and more of a basis for the work.

c. Of course, if sufficient of the key text is reconstructed, the cryptanalyst might identify the book that is being used for the key, and if available, his subsequent labors are very much simplified.

d. All the foregoing is, however, dependent upon having a knowledge of the primary components or cipher alphabets employed in the encipherment. Even if the primary components are differently-mixed sequences, so long as they are known sequences, the procedure is quite obvious in view of the foregoing explanation. The training the student has already had is believed sufficient to indicate to him the procedure he may follow in that solution, and no further details will here be given in respect to such cases. But what if the primary components are not known sequences? This contingency will be treated presently.

39. Solution of a progressive-alphabet cipher when the cipher alphabets are known. - a. Taking a very simple case, suppose the interacting elements referred to in Par. 36b consist merely of two primary cipher components which slide against each other to produce a set of 26 secondary cipher alphabets.

Beginning at an initial juxtaposition, producing say, alphabet 1, the subsequent secondary alphabets are in the sequence 2, 3, ... 26, 1, 2, 3 ... , and so on. If a different initial juxtaposition is used, say alphabet 10 is the first one, the sequence is exactly the same as before, only beginning at a different point.

b. Suppose the two primary components are based upon the keyword

HYDRAULIC. A message is to be enciphered, beginning with alphabet 1. Thus:

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

Letter No...	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Alphabet ...	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Plain text .	E	N	E	M	Y	H	A	S	P	L	A	C	E	D	H	E	A	V	Y	I	N	T	E	R
Cipher text.	E	O	G	P	U	U	E	Y	H	M	K	Q	V	M	K	Z	S	J	Q	H	E	N	L	H

Letter No...	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
Alphabet ...	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Plain text .	D	I	C	T	I	O	N	F	I	R	E	U	P	O	N	Z	A	N	E	S
Cipher text.	H	L	C	V	B	S	S	N	J	E	P	K	D	D	D	G	P	U	H	F

Letter No...	45	46	47	48	49	50	51	52	53
Alphabet ...	19	20	21	22	23	24	25	26	1
Plain text .	V	I	L	L	E	R	O	A	D
Cipher text.	K	H	H	Y	L	H	M	R	D

c. This method reduces to a periodic system involving 26 secondary cipher alphabets and the latter are used in simple progression. It is obvious therefore that the 1st, 27th, 53d, ... letters are in the 1st alphabet; the 2d, 28th, 54th, ... letters are in the 2d alphabet, and so on.

d. To solve such a cryptogram, knowing the two primary components, is hardly a problem at all. The only element lacking is a knowledge of the starting point. But this is not necessary, for merely by completing the plain-component sequences and examining the diagonals of the diagram, the plain text becomes evident. For example, given the following: H I D C T E H U X I. Completing the plain-component sequences initiated by the successive cipher letters, the plain text, E N E M Y M A C H I... is seen to come out in

successive steps upward in Fig. 24. Had the cipher component been shifted in the opposite direction in encipherment, the steps would have been downward instead of upward. If the sliding strips are set up according to the sequence of cipher letters but on a diagonal, then of course, the plain-text letters would have reappeared on one generatrix.

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

e. The student will understand what simple

Figure 24.

modifications in procedure would be required in case the two primary components were different mixed sequences. But what if the primary components are not known sequences? How does the cryptanalyst proceed in that case?

40. General solution for ciphers involving a long keying sequence of fixed length and composition. - a. It is obvious, as stated at a previous point, that no matter how the keying sequence is derived, if all the communicants employ the same key, or if this key is used many times by a single office, and if it always begins at the same point, the various messages may simply be superimposed. Thus, their respective 1st, 2d, 3rd, ... letters will all fall within columns which have been enciphered by the 1st, 2d, 3rd, ... key letters. If there is a sufficient number of messages, solution then becomes possible by frequency analysis of the successive columns -- no matter how long the keying sequence may be. This method of solution by superimposition has already been outlined in Par. 20 and no further reference to it need here be made.

b. But now suppose that the keying sequence does not always begin at the same point for all messages. Suppose the several communicants are able to select at will any element of the keying sequence and employ it as the initial keyletter. Thus, such a keying sequence, if regarded as partaking of

the nature of a circle, will afford as many possible starting points as there are letters or characters in that sequence. Now if there are no external indications or indicators in the cryptograms pertaining to such a system, such as would afford enemy cryptanalysts direct and definite information with regard to the initial keying element for each cryptogram, then it would seem as though the superimposition of messages (to bring letters enciphered by the same cipher alphabets within the same columns) would be difficult or impossible, and therefore that attempts at solution are blocked at their very beginning. This, however, is not the end of the story. For suppose two of the messages have in common only one polygraph, say of 5 letters; these two messages may be juxtaposed so as to bring these repetitions into superimposition. Thus, the possession of this long polygraph in common serves to "tie" these two messages together or to "interlock" them. Then, suppose a shorter polygraph, say of 4 letters, is possessed in common by one of these two messages and a third message; this will serve to tie in the latter with the first two. Extension of this process, including the data from shorter repetitions of trigraphs and digraphs, will serve to assemble a whole set of such messages in proper superimposition. Therefore, the first step is to examine all the messages for repetitions.

c. When such repetitions are found, and if there are plenty of them so that assumptions for probable words are easy to make, it is clear that the correct assumptions will enable the cryptanalyst to set up plain-cipher equivalencies which will make it possible to reconstruct the primary components. Depending upon the type used, the principles of direct or indirect symmetry of position will be very useful in this process.

d. But if it happens that there are no polygraphs by means of which two or more messages may be tied together and properly superimposed, the simple

methods mentioned in subparagraphs a-c cannot here be applied. However, although the road toward a solution seems to be blocked rather effectively, there is a detour which presents rather interesting vistas. The latter are really of such importance in cryptanalysis as to warrant detailed treatment.

SECTION XI

THE CRYPTANALYTIC COINCIDENCE TEST.

	Paragraph
The basic theory of the coincidence test	41
General procedure to be followed in making the K -test	42
Example of application of the K -test	43
Subsequent steps	44

41. The basic theory of the coincidence or K -test. - a. In Appendix 2 of the preceding text¹ certain simple applications of the theory of probability were presented for the student's consideration, by way of pointing out to him the important role which certain phases of that branch of mathematics play in cryptanalysis. Reference was there made to the subject of coincidences and the bearing it has in connection with the study of repetitions in cryptograms. In this section the matter will be pursued a few steps further.

b. In the appendix referred to, it was shown that the probability of monographic coincidence (1) in random text employing a 26-letter alphabet is .0385; (2) in English telegraphic plain text, .0667. These two parameters were represented by the symbols K_r and K_p , respectively. The important role which these values play in a certain cryptanalytic test will now be explained.

c. One of the most important techniques in cryptanalytics is that known as applying the coincidence or "kappa" test. This test is useful for several

¹Military Cryptanalysis, Part II. It is recommended that the student refresh his memory by reviewing this appendix.

- 102 -

cryptanalytic purposes and one of the most important of them is to ascertain when two or more sequences of letters are correctly superimposed. By the word "correct" in this case is merely meant that the sequences are so arranged relatively as to facilitate or make possible a solution. The test has for its theoretical basis the following circumstances:

(1) If any two rather lengthy sequences of characters are superimposed, it will be found, on examining both members of the successive pairs of letters brought into vertical juxtaposition, that in a certain number of cases the two superimposed letters will coincide.

(2) If both sequences of letters constitute random text (of a 26-letter alphabet), there will be about 38 or 39 such cases of coincidence per thousand pairs examined. This, of course, is because $K_r = .0385$.

(3) If both sequences of letters constitute plain text, there will be about 66 or 67 such cases of coincidence per thousand pairs examined. This is because K_p is .0667.

(4) If the superimposed sequences are wholly monoalphabetic encipherments of plain text by the same cipher alphabet, there will still be about 66 or 67 cases of coincidence in each 1000 cases examined, because in monoalphabetic substitution there is a fixed or unvarying relation between plaintext letters and cipher letters so that for statistical purposes monoalphabetic cipher text behaves just the same as if it were normal plain text.

(5) Even if the two superimposed sequences are not monoalphabetically enciphered texts, but are polyalphabetic in character, there will still be about 66 or 67 cases of identity between superimposed letters per thousand cases examined, provided the two sequences really belong to the same cryptographic system and are superimposed at the proper point with respect to the keying sequence. The reasons for this will be set forth in the succeeding

sub-paragraphs.

(6) Consider the two messages below. They have been enciphered polyalphabetically by the same two primary components sliding against each other. The two messages use the same keying sequence, beginning at the same initial point in that sequence. Consequently, the two messages are identically enciphered, letter for letter, and the only differences between them are those occasioned by differences in plain text.

	{	Alphabets...	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	12	6	
No.1.	{	Plain text..	W	H	E	N	I	N	T	H	E	C	O	U	R	S	E	L	O	N	G	M	...
	{	Cipher.....	E	Q	N	B	T	F	Y	R	C	X	X	L	Q	J	N	Z	O	Y	A	W	...

	{	Alphabets...	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	12	6	...
No.2.	{	Plain text..	T	H	E	G	E	N	E	R	A	L	A	B	S	O	L	U	T	E	L	Y	...
	{	Cipher.....	P	Q	N	T	U	F	B	W	D	J	L	Q	H	Y	Z	P	T	M	Q	I	...

Note, now, that (a) in every case in which two superimposed cipher letters are the same, the plain-text letters are identical and (b) in every case in which two superimposed cipher letters are different, the plain-text letters are different. In such a system, even though the cipher alphabet changes from letter to letter, the number of cases of identity or coincidence in the two members of a pair of superimposed cipher letters will still be about 66 or 67 per thousand cases examined, because the two members of each pair of superimposed letters are in the same cipher alphabet and it has been seen in (4) that in monoalphabetic cipher text K is the same as for plain text², viz, .0667. The two messages may here be said to be superimposed "correctly", that is, brought into proper juxtaposition with respect to the keying sequence.

(7) But now suppose the same two messages are superimposed "incorrectly", that is, they are no longer in proper juxtaposition with res-

² The fact that in this case each monoalphabet contains but two letters does not affect the theoretical value of K; and whether the actual number of coincidences agrees closely with the expected number based upon $K = .0667$ depends upon the lengths of the two superimposed sequences.

pect to the keying sequence. Thus:

	Alphabets...	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	12	...
No.1	Plain text..	W	H	E	N	I	N	T	H	E	C	O	U	R	S	E	L	O	N	G	...
	Cipher.....	E	Q	N	B	<u>T</u>	F	Y	R	C	X	X	<u>L</u>	Q	J	N	Z	O	Y	A	...

	Alphabets...	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	...
	Plain text..	T	H	E	G	E	N	E	R	A	L	A	B	S	O	L	U	T	E	...
	Cipher.....	P	Q	N	<u>T</u>	U	F	B	W	D	J	<u>L</u>	Q	H	Y	Z	P	T	M	...

It is evident that the two members of every pair of superimposed letters are no longer in the same cipher alphabet, and therefore, if two superimposed cipher letters are identical this is merely an "accident", for now there is no basic or general cause for the similarity, such as is true in the case of a correct superimposition. The similarity, if present, is, as already stated, due to chance and the number of such cases of similarity should be about the same as though the two cipher letters were drawn at random from random text, in which $K_r = .0385$. It is no longer true that (a) in every case in which two superimposed cipher letters are the same, the plain-text letters are identical, or (b) in every case in which two superimposed cipher letters are different, the plain-text letters are different. Note, for example, that the superimposed T_c 's represent two different plain-text letters and that the S_p of the word COURSE in the 1st message gives J_c while the S of the word ABSOLUTELY in the 2d message gives H_c . Thus, it becomes clear that in an incorrect superimposition two different plain-text letters enciphered by two different alphabets may "by chance" produce identical cipher letters, which on superimposition yield a coincidence having no external indications as to dissimilarity in plain-text equivalents. Hence, if there are no other factors which enter into the matter and which might operate to distort the results to be expected from the operation of the basic factor, the expected number of cases of identical cipher letters brought together by an incorrect superimposition will be determined by the value $K_r = .0385$.

(8) But now note also that in the foregoing incorrect superimposition there are two Z_c 's and that they represent the same plain-text letter L. This is occasioned by the fact that the plain-text messages happened to have L's in just those two places and that the cipher alphabet happened to be the same both times. Hence, it becomes clear that the same cipher alphabet brought into play twice may "by chance" happen to encipher the same plain-text letter both times, thus producing identical cipher letters. In some systems this source of identity in superimposed cipher letters is of little importance, in other systems it may materially affect the actual number of coincidences. For instance, if a system is such that it produces a long secondary keying cycle composed of repetitions of short primary keying cycles, an incorrect superimposition of two cryptograms may bring into juxtaposition many of these short cycles, with the result that the actual number of cases of identical superimposed cipher letters is much greater than the expected number based upon $K_r = .0385$. Thus, this source for the production of identical cipher letters in an incorrect superimposition operates to increase the number of cases to be expected from the fundamental constant $K_r = .0385$.

(9) In some systems, where nonrelated cipher alphabets are employed, it may happen that two identical plain-text letters may be enciphered by two different cipher alphabets which, "by chance", have the same equivalent for the plain-text letter concerned. This is, however, a function of the particular cryptographic system and can be taken into account when the nature of the system is known.

(10) In general, then, it may be said that in the case of a correct superimposition the probability of identity or coincidence in superimposed cipher letters is .0667; in the case of an incorrect superimposition, the probability is at least .0385 and may be somewhat greater, depending upon

special circumstances. The foregoing situation and facts make possible what has been referred to as the "coincidence test." Since this test uses the constant K , it is also called the "kappa test."

d. The way in which the coincidence test may be applied will now be explained. The statement that $K_p = .0667$ means that in 1000 cases where two letters are drawn at random from a large volume of plain text, there will be about 66 or 67 cases in which the two letters coincide, that is, are identical. Nothing is specified as to what the two letters shall be; they may be two Z's or they may be two E's. This constant, .0667, really denotes a percentage: if many comparisons of single letters are made, the letters being drawn at random from among those constituting a large volume of plain text, 6.67% of these comparisons made will yield coincidences. So, if 2000 such comparisons are made, the theory indicates that there should be about $.0667 \times 2000 = 133$ coincidences; if there is sufficient text to permit of making 20,000 comparisons, there should be about 1334 coincidences, and so on.

e. Another way of handling the matter is to find the ratio of the observed number of coincidences to the total number of cases in which the event in question might possibly occur, i.e., the total number of comparisons of superimposed letters. When this ratio is closer to .0667 than it is to .0385 the correct superimposition has been ascertained. This is true because in the case of a correct superimposition both members of each pair of superimposed letters actually belong to the same monoalphabet and therefore the probability of their coinciding is .0667; whereas in the case of an incorrect superimposition the members of each pair of superimposed letters belong, as a general rule, to different monoalphabets³, and therefore the probability of

³The qualifying phrase "as a general rule" is intended to cover any distortion in results occasioned by the presence of an unusual number of those cases of coincidence described under subparagraph c(8) and (9).

their coinciding is nearer .0385 than .0667.

f. From the foregoing, it becomes clear that ^{the} /kappa-test involves ascertaining the total number of comparisons that can be made in a given case, as well as ascertaining the actual number of coincidences in the case under consideration. When only two messages are superimposed, this is easy: the total number of comparisons that can be made is the same as the number of superimposed pairs of letters. But when more than two messages are superimposed in a superimposition diagram it is necessary to make a simple calculation, based upon the fact that n letters yield $\frac{n(n-1)}{2}$ pairs or comparisons, where n is the number of letters in the column⁴. For example, in the case of a column of 3 letters, there are $\frac{3 \times 2}{2} = 3$ comparisons. This can be checked by noting that the 1st letter in the column may be compared with the 2d, the 2d with the 3rd, and the 1st with the 3rd, making 3 comparisons in all. The number of comparisons per column times the number of columns in the superimposition diagram of letters gives the total number of comparisons. The extension of this reasoning to the case where a superimposition diagram has columns of various lengths is quite obvious: one merely adds together the number of comparisons for columns of different lengths to obtain a grand total. For convenience, the following brief table is given:

⁴This has already been encountered (Footnote 3, Appendix 2, Military Crypt-analysis, Part 2.) It is merely a special case under the general formula for ascertaining the number of combinations that may be made of n different things taken r at a time, which is $C_r^n = \frac{n!}{r!(n-r)!}$. In studying coincidences by the method indicated, since only two letters are compared at a time, r is always 2; hence the expression $\frac{n!}{r!(n-r)!}$, which is the same as $\frac{n(n-1)(n-2)!}{2(n-2)!}$, becomes by cancellation of $(n-2)!$, reduced to $\frac{n(n-1)}{2}$.

No. of letters in column	No. of comparisons	No. of letters in column	No. of comparisons	No. of letters in column	No. of comparisons
2	1	11	55	21	210
3	3	12	66	22	231
4	6	13	78	23	253
5	10	14	91	24	276
6	15	15	105	25	300
7	21	16	120	26	325
8	28	17	136	27	351
9	36	18	153	28	378
10	45	19	171	29	406
		20	190	30	435

g. In ascertaining the number of coincidences in the case of a column containing several letters, it is again necessary to use the formula $\frac{n(n-1)}{2}$, only in this case n is the number of identical letters in the column. The reasoning, of course, is the same as before. The total number of coincidences is the sum of the number of coincidences for each case of identity. For example, in the column shown at the side, containing 10 letters, there are

C 3 B's, 2 C's, 4 K's, and one Z. The 3 B's yield 3 coincidences, the
 K
 B 2 C's yield one coincidence, and the 4 K's yield 6 coincidences. The
 K
 Z sum of $3+1+6$ makes a total of 10 coincidences in 45 comparisons.

K
 C 42. General procedure to be followed in making the K-test. - a. The
 B
 B steps in applying the foregoing principles to an actual case will now
 K
 be described. Suppose several messages enciphered by the same keying sequence but each beginning at a different point in that sequence are to be solved. The indicated method of solution is that of superimposition, the problem being to determine just where the respective messages are to be superimposed so that the cipher text within the respective columns formed by the superimposed messages will be monoalphabetic. From what has been indicated above, it will be understood that the various messages may be shifted relative to one another to many different points of superimposition, there being but one correct superimposition for each message with respect to all the others.

First, all the messages are numbered according to their lengths, the longest being assigned the number 1. Commencing with messages 1 and 2, and keeping number 1 in a fixed position, message 2 is placed under it so that the initial letters of the two messages coincide. Then the two letters forming the successive pairs of superimposed letters are examined and the total number of cases in which the superimposed letters are identical is noted, this giving the observed number of coincidences. Next, the total number of superimposed pairs is ascertained, and the latter is multiplied by .0667 to find the expected number of coincidences. If the observed number of coincidences is considerably below the expected number, or if the ratio of the observed number of coincidences to the total number of comparisons is nearer .0385 than .0667, the superimposition is incorrect and message 2 is shifted to the next superimposition, that is, so that its 1st letter is under the 2d of message 1. Again the observed number of coincidences is ascertained and is compared with the expected number. Thus, by shifting message 2 one space at a time (to the right or left relative to message 1) the coincidence test finally should indicate the proper relative positions of the two messages. When the correct point of superimposition is reached the cryptanalyst is rarely left in doubt, for the results are sometimes quite startling. After messages 1 and 2 have been properly superimposed message 3 is tested first against messages 1 and 2 separately, and then against the same two messages combined at their correct superimposition.⁵ Thus message 3 is shifted a step each time until its correct position with respect to messages 1 and 2 has been found. Then message 4 is taken and its proper point of superimposition with respect to messages 1, 2,

⁵ At first thought the student might wonder why it is advisable or necessary to test message 3 against message 1 and message 2 separately before testing it against the combination of messages 1 and 2. The first two tests, it seems to him, might be omitted and time saved thereby. The matter will be explained in Par. 43f(3).

- 110 -

and 3 is ascertained. The process is continued in this manner until the correct points of superimposition for all the messages have been found. It is obvious that as messages are added to the superimposition diagram, the determination of correct points of superimposition for subsequent messages becomes progressively more certain and therefore easier.

b. In the foregoing procedure it is noted that there is necessity for repeated displacement of one message against another or other messages. Therefore, it is advisable to transcribe the messages on long strips of cross-section paper, joining sections accurately if several such strips are necessary to accommodate a long message. Thus, a message once so transcribed can be shifted to various points of superimposition relative to another such message, without repeatedly rewriting the messages.

c. Machinery for automatically comparing letters in applying the coincidence test has been devised. Such machines greatly facilitate and speed up the procedure.

43. Example of application of the K-test. - a. With the foregoing in mind, a practical example will now be given. The following ^{messages,} assumed to be the first 4 of a series of 30 messages, supposedly enciphered by a long keying sequence, but each message commencing at a different point in that sequence, are to be arranged so as to bring them into correct superimposition:

Message 1.

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

- 111 -

Message 2.

CWHPK KXFLU MKURY XCOPY WNJUW KWIHL
 OKZTL AWRDF GDDEZ DLBOT FUZNA SRHHJ
 NGUZK PRC DK YOOBV DDXCD OGRGI RMICN
 HSGGO PYAOY X

Message 3.

WFWTD NHTGM RAAZG PJDSQ AUPFR OXJRO
 HRZWC ZSRTE EEVPX OATDQ LDOQZ HAWNX
 THDXL HYIGK VYZWX BKOQO AZQND TNALT
 CNYEH TSCT

Message 4.

TULDH NQEZ Z UTYGD UEDUP SDLIO LNNBO
 NYLQQ VQGCD UTUBQ XSOSK NOXUV KCYJX
 CNJKS ANGUI FTOWO MSNBQ DBAIV IKNWG
 VSHIE P

b. Superimposing ⁶ messages 1 and 2, beginning with their 1st letters,

No. 1 ... PGLPNHUF³³FRKSAUQQAQYUOZAKGAEQQCNPRK
 No. 2 ... CWHPK³³KKXFLUMKURYXCOPHWNJUWKWIHLKZ

No. 1 ... OVHYEIUYNBONNFDMWZLUKQAQAHZMGCDSL⁶⁶
 No. 2 ... TLAWRDFGDDEZDLBOTFUZNASRHHJNGUZKP

No. 1 ... EAGCJPIVJWVAUDBAHMIHKORMLTFYZLGSOGK⁹⁹
 No. 2 ... RCDKYOOBVDDXCDOGRGI⁹⁹RMICNHSGGOPYAOYX

the number of coincidences is found to be 7. Since the total number of comparisons is 101, the expected number, if the superimposition were correct, should be $101 \times .0667 = 6.7367$, or about 7 coincidences. The fact that the observed number of coincidences matches the expected number on the very first trial

⁶The student will have to imagine the messages written out as continuous sequences on cross-section paper.

- 112 -

creates an element of suspicion: such good fortune is rarely the lot of the practical cryptanalyst. It is very unwise to stop at the first trial, even if the results are favorable, for this close agreement between theoretical and actual numbers of coincidences might just be "one of those accidents." Therefore message 2 is shifted one space to the right, placing its 1st letter beneath the 2d letter of message 1. Again the number of coincidences is noted and this time it is found to be only 4. The total number of comparisons is now 100; the expected number is still about 7. Here the observed number of coincidences is considerably less than the expected number, and when the relatively small number of comparisons is borne in mind, the discrepancy between the theoretical and actual results is all the more striking. The hasty cryptanalyst might therefore jump to the conclusion that the 1st superimposition is actually the correct one. But only two trials have been made thus far and a few more are still advisable, for in this scheme of superimposing a series of messages it is absolutely essential that the very first superimpositions rest upon a perfectly sound foundation -- otherwise subsequent work will be very difficult, if not entirely fruitless. Additional trials will therefore be made.

c. Message 2 is shifted one more space to the right and the number of coincidences is now found to be only 3. Once again message 2 is shifted, to the position shown below, and the observed number of coincidences jumps suddenly to 9.

No. 1 ...P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K
 No. 2 ... C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L

No. 1 ...O V H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L
 No. 2 ...O K Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U

No. 1 ...E A G C J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K
 No. 2 ...Z K P R C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A

No. 1 ...
 No. 2 ...O Y X

The total number of comparisons is now 98, so that the expected number of coincidences is $98 \times .0667 = 6.5366$, or still about 7. The 2d and 3rd superimpositions are definitely incorrect; as to the 1st and 4th, the latter gives almost 30% more coincidences than the former. Again considering the relatively small number of comparisons, this 30% difference in favor of the 4th superimposition as against the 1st is important. Further detailed explanation is unnecessary, and the student may now be told that it happens that the 4th superimposition is really correct; if the messages were longer, all doubt would be dissolved. The relatively large number of coincidences found at the 1st superimposition is purely accidental in this case.

d. The phenomenon noted above, wherein the observed number of coincidences shows a sudden increase in moving from an incorrect to a correct superimposition is not at all unusual, nor should it be unexpected, because there is only one correct superimposition, while all other superimpositions are entirely incorrect. In other words, a superimposition is either 100% correct or 100 % wrong -- and there are no gradations between these two extremes. Theoretically, therefore, the difference between the correct superimposition and any one of the many incorrect superimpositions should be very marked, since it follows from what has been noted above, that one cannot expect that the discrepancy between the actual and the theoretical number of coincidences should get smaller and smaller as one approaches closer and closer to the correct superimposition.⁷ For if letters belonging to the same cipher alphabet are regarded as being members of the same family, so to speak, then the two letters forming the successive pairs of letters brought into superimposi-

⁷The importance of this remark will be appreciated when the student comes to study longer examples, in which statistical expectations have a better opportunity to come into play.

- 114 -

tion by an incorrect placement of one message relative to another are total strangers to each other, brought together by pure chance. This happens time and again, as one message is slid against the other -- until the correct superimposition is reached, whereupon in every case the two superimposed letters belong to the same family. There may be many different families (cipher alphabets) but the fact that in every case there are two members of the same family represented causes the marked jump in number of coincidences.

e. In shifting one message against another, the cryptanalyst may move to the right constantly, or he may move to the left constantly, or he may move alternately to the left and right from a selected initial point. Perhaps the latter is the best plan.

f. (1) Having properly superimposed messages 1 and 2, message 3 is next to be studied. Now it is of course possible to test the latter message against the combination of the former, without further ado. That is, ascertaining merely the total number of coincidences given by the superimposition of the 3 messages might be thought sufficient. But for reasons which will soon become apparent it is better, even though much more work is involved, first to test message 3 against message 1 alone and against message 2 alone. This will really not involve much additional work after all, since the two tests can be conducted simultaneously, because the proper superimposition of messages 1 and 2 is already known. If the tests against messages 1 and 2 separately at a given superimposition give good results, then message 3 can be tested, at that superimposition, against messages 1 and 2 combined. That is, all 3 messages are tested as a single set. Since, according to the scheme outlined, a set of three closely related tests is involved, one might as well systematize the work so as to save time and effort, if possible. With this in view a diagram such as that at the side (page 115) is made and in it the

	1	2	3
1	X		
2	X	X	
3	X	X	X

The successive columns are now examined and the coincidences are recorded, remembering that only coincidences between messages 1 and 3, and between messages 2 and 3 are now to be tabulated in the diagram. The results for this first test are shown in the diagram at the side. This superimposition yields but 3 coincidences between messages 1 and 3, and the same number between messages 2 and 3. The total numbers of comparisons are then noted and the following table is drawn up:

Combination	Total No. of Comparisons	No. of Coincidences		Discrepancy
		Expected	Observed	
Messages 1 and 3	99	about 7	3	-57%
Messages 2 and 3	96	" 6	3	-50%
Messages 1, 2, and 3	291	" 19	15	-21%

(3) The reason for the separate tabulation of coincidences between messages 1 and 3, 2 and 3, and 1, 2, and 3 should now be apparent. Whereas the observed number of coincidences is 57% below the expected number of coincidences in the case of messages 1 and 3 alone, and 50% below in the case of messages 2 and 3 alone, the discrepancy between the expected and observed numbers is not quite so marked (-21%) when all three messages are considered together, because the relatively high number of coincidences between messages 1 and 2, which are correctly superimposed, serves to counterbalance the low numbers of coincidences between 1 and 3, and 2 and 3. Thus, a correct superimposition for one of the three combinations may yield such good results as to mask the bad results for the other two combinations.

(4) Message 3 is then shifted one space to the right, and the same procedure is followed as before. The results are shown below:

No. 1 ... P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K O V ³⁵
 No. 2 ... C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L O K
 No. 3 ... W F W T D N H T G M R A A Z G P J D S Q A U P F R O X J R O H R Z W
 No. 1 ... H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L E A G C ⁷
 No. 2 ... Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U Z K P F
 No. 3 ... C Z S R T E E E V P X O A T D Q L D O Q Z H A W N X T H D X L H Y I C
 No. 1 ... J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K ¹⁰¹
 No. 2 ... C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A O Y X
 No. 3 ... K V Y Z W X B K O Q O A Z Q N D T N A L T C N Y E H T S C T

	1	2	3
1	X	 	
2	X	X	
3	X	X	X

Combination	Total No. of Comparisons	No. of Coincidences		Discrepancy
		Expected	Observed	
Messages 1 and 3	99	about 7	10	+43%
Messages 2 and 3	97	" 6	6	0%
Messages 1, 2 and 3	293	" 20	25	+25%

Note how well the observed and expected numbers of coincidences agree in all three combinations. Indeed, the results of this test are so good that the cryptanalyst might well hesitate to make any more tests.

(5) Having ascertained the relative positions of 3 messages, the fourth message is now studied. Here are the results for the correct superimposition.

No. 1 ... P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K O ³¹
 No. 2 ... C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L O
 No. 3 ... W F W T D N H T G M R A A Z G P J D S Q A U P F R O X J R O H R Z
 No. 4 ... T U L D H N Q E Z Z U T Y G D U E D U P S D L I Q L N N B O N Y
 No. 1 ... V H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L E A ⁶⁸
 No. 2 ... K Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U Z K
 No. 3 ... W C Z S R T E E E V P X O A T D Q L D O Q Z H A W N X T H D X L H Y
 No. 4 ... L Q Q V Q G C D U T U B Q X S O S K N Q X U V K C Y J X C N J K S A
 No. 1 ... G C J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K ¹⁰¹
 No. 2 ... P R C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A O Y X
 No. 3 ... I G K V Y Z W X B K O Q O A Z Q N D T N A L T C N Y E H T S C T
 No. 4 ... N G U I F T O W O M S N B Q D B A I V I K N W G V S H I E P

	1	2	3	4	Combination	Total No. of Comparisons	No. of Coincidences		Discrepancy
							Expected	Observed	
1	X	Z	Z	Z					
2	X	X	Z	Z	Messages 1 and 4	96	about 6	7	+16%
					Messages 2 and 4	95	" 6	6	0%
3	X	X	X	Z	Messages 3 and 4	96	" 6	5	-16%
4	X	X	X	X	Messages 1, 2, 3 and 4	581	" 39	44	+13%

The results for an incorrect superimposition (1st letter of message 4 under 4th letter of message 1) are also shown for comparison:

No. 1 ... P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K O
 No. 2 ... C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L O
 No. 3 ... W F W T D N H T G M R A A Z G P J D S Q A U P F R O X J R O H R Z
 No. 4 ... T U L D H N Q E Z Z U T Y G D U E D U P S D L I O L N N B O N

No. 1 ... V H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L E A
 No. 2 ... K Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U Z K
 No. 3 ... W C Z S R T E E E V P X O A T D Q L D O Q Z H A W N X T H D X L H Y
 No. 4 ... Y L Q Q V Q G C D U T U B Q X S O S K N O X U V K C Y J X C N J K S

No. 1 ... G C J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K
 No. 2 ... P R C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A O Y X
 No. 3 ... I G K V Y Z W X B K O Q O A Z Q N D T N A L T C N Y E H T S C T
 No. 4 ... A N G U I F T O W O M S N B Q D B A I V I K N W G V S H I E P

	1	2	3	4	Combination	Total No. of Comparisons	No. of Coincidences		Discrepancy
							Expected	Observed	
1	X	Z	Z	"					
2	X	X	Z	"	Messages 1 and 4	96	about 6	2	-67%
					Messages 2 and 4	96	" 6	3	-50%
3	X	X	X	"	Messages 3 and 4	96	" 6	2	-67%
4	X	X	X	X	Messages 1, 2, 3 and 4	582	" 39	33	-15%

(6) It is believed that the procedure has been explained with sufficient detail to make further examples unnecessary. The student should bear in mind always that as he adds messages to the superimposition diagram it is necessary that he recalculate the number of comparisons so that the correct expected or theoretical number of coincidences will be before him to compare with the observed number. In adding messages he should see that the results of the separate tests are consistent, as well as those for the combined tests, otherwise he may be led astray at times by the overbalancing effect of the large number of coincidences for the already ascertained, correct superimpositions.

44. Subsequent steps. - a. In Par. 43a four messages were given of a series supposedly enciphered by a long keying sequence, and the succeeding paragraphs were devoted to an explanation of the preparatory steps in the solution. The messages have now been properly superimposed, so that the text has been reduced to monoalphabetic columnar form, and the matter is now to be pursued to its ultimate stages.

b. The four messages employed in the demonstration of the principles of the K-test have served their purpose. The information that they are messages enciphered by an intelligible running key, by reversed standard cipher alphabets, was withheld from the student, for pedagogical reasons. Were the key a random sequence of letters instead of intelligible text, the explanation of the coincidence test would have been unchanged in the slightest particular, so far as concerns the mechanics of the text itself. Were the cipher alphabets unknown, mixed alphabets, the explanation of the K-test would also have been unchanged in the slightest particular. But, as stated before, the four messages actually represent encipherments by means of an intelligible running key, by reversed standard alphabets; they will now be used to illustrate the solution

of cases of this sort.

c. Assuming now that the cryptanalyst is fully aware that the enemy is using the running key system with reversed standard alphabets (obsolete U. S. Army Cipher Disk), the method of solution outlined in Par. 38 will be illustrated, employing the first of the four messages referred to above, that beginning PGLPN HUF^RK SAUQQ. The word DIVISION will be taken as a probable word and tested against the key, beginning with the very first letter of the message. Thus:

Cipher text	P G L P N H U F R K S A U Q Q ...
Assumed text	D I V I S I O N
Resultant key text ...	S O G X F ...

The resultant key text is unintelligible and the word DIVISION is shifted one letter to the right.

Cipher text	P G L P N H U F R K S A U Q Q ...
Assumed text	D I V I S I O N
Resultant key text ...	J T K . . .

Again the resultant key text is unintelligible and the hypothetical word DIVISION is shifted once more. Continuation of this process to the end of the message proves that the word is not present. Another probable word is assumed: REGIMENT. When the point shown below is reached, note the results:

Cipher text	P G L P N H U F R K S A U Q Q
Assumed text	R E G I M E N T
Resultant key text ...	E L A N D O F T

It certainly looks as though intelligible text were being obtained as key text.

The words LAND OF T... suggest that THE be tried. The keyletters HE give NO, making the plain text readREGIMENT NO... . The four spaces preceding REGIMENT suggest such words as HAVE, SEND, MOVE, THIS, etc. A clue may be found by assuming that the E before LAND in the key is part of the word THE. Testing it on the cipher text gives IS for the plain text, which certainly indicates that the message begins with the word THIS. The latter yields IN

- 121 -

for the first two key letters. And so on, the process of checking one text against the other continuing until the entire message and the key text have been reconstructed.

d. Thus far the demonstration has employed but one of the four messages available for solution. When the reconstruction process is applied to all four simultaneously it naturally goes much faster, with reduced necessity for assuming words after an initial entering wedge has been driven into one message. For example, note what happens in this case; just as soon as the word REGIMENT is tried in the proper place:

Key text E L A N D O F T																
No. 1	{ Cipher text ...	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	. . .	
	{ Plain text	R	E	G	I	M	E	N	T							
No. 2	{ Cipher text ...	---	C	W	H	P	K	K	X	F	L	U	M	K			
	{ Plain text	I	E	L	D	T	R	A	I							
No. 3	{ Cipher text ...	---	W	F	W	T	D	N	H	T	G	M	R	A	A	Z	. . .	
	{ Plain text	L	I	N	G	K	I	T	C							
No. 4	{ Cipher text ...	---	T	U	L	D	H	N	Q	E	Z	Z	U	T	Y	. . .		
	{ Plain text	T	I	T	A	N	K	G	U							

It is obvious that No. 2 begins with FIELD TRAIN; No. 3, with ROLLING KITCHEN; No. 4 with ANTITANK GUN. These words yield additional keyletters, the latter suggest additional plain text, and thus the process goes on until the solution is completed.

e. But now suppose that the key text that has been actually employed in encipherment is not intelligible text. The process is still somewhat the same only in this case one must have at least two messages in the same key. For instead of checking a hypothetical word (assumed to be present in one message) against the key, the same kind of a check is made against the other message or messages. Assume, for instance, that in the case just described the key text, instead of being intelligible text, were a series of letters .

produced by applying a rather complex transposition to an originally intelligible key text. Then if the word REGIMENT were assumed to be present in the proper place in message No. 1 the resultant keyletters would yield an unintelligible sequence. But these keyletters when applied to message No. 2 would nevertheless yield IELDTRAIN; when applied to message No. 3, LINGKITC, and so on. In short, the text of one message is checked against the text of another message or messages; if the originally assumed word is correct, then plain text will be found in the other messages.

f. All the foregoing work is, of course, based upon a knowledge of the cipher alphabets employed in the encipherment. What if the latter are unknown sequences? It may be stated at once that not much could be done with but four messages, even after they had been superimposed correctly, for the most that one would have in the way of data for the solution of the individual columns of text would be four letters per alphabet -- which is not nearly enough. Data for solution by indirect symmetry by the detection of isomorphs cannot be expected, for no isomorphs are produced in this system. Solution can be reached only if there is sufficient text to permit of the analysis of the columns of the superimposition diagram. When there is this amount of text there are also repetitions which afford bases for the assumption of probable words. Only then, and after the values of a few cipher letters have been established can indirect symmetry be applied to facilitate the reconstruction of the primary components - if used.

g. Even when the volume of text is great enough so that each column contains say 15 to 20 letters, the problem is still not an easy one. But frequency distributions with 15 to 20 letters can usually be studied statistically, so that if two distributions present similar characteristics, the latter may be used as a basis for combining distributions which pertain to the same

cipher alphabet. The next section will be devoted to a detailed treatment of the implications of the last statement.

SECTION XII

THE MATCHING OF FREQUENCY DISTRIBUTIONS.

	Paragraph
Preliminary remarks	45
The "Cross-product" or "X-test" in cryptanalysis.....	46
Derivation of the X-test.....	47
Applying the X-test in matching distributions.....	48
Practical example of applying the X-test.....	49
The " Φ (Phi)-test".....	50
Derivation of the Φ -test.....	51
Applying the Φ -test.....	52

45. Preliminary remarks. - a. The real purpose of making the coincidence test in cases such as that studied in the preceding section is to permit the cryptanalyst to arrange his data so as to circumvent the obstacle which the enemy, by adopting a complicated polyalphabetic scheme of encipherment, places in the way of solution. The essence of the matter is that by dealing individually with the respective columns of the superimposition diagram the cryptanalyst has arranged the polyalphabetic text so that it can be handled as though it were monoalphabetic. Usually, the solution of the latter is a relatively easy matter, especially if there is sufficient text in the columns, or if the letters within certain columns can be combined into single frequency distributions, or if some cryptographic relationship can be established between the columns.

b. It is obvious that merely ascertaining the correct relative positions of the separate messages of a series of messages in a superimposition diagram is only a means to an end, and not an end in itself. The purpose is, as already stated, to reduce the complex, heterogeneous, polyalphabetic text to simple, homogeneous, monoalphabetic text. But the latter can be solved only when there

are sufficient data for the purpose -- and that depends often upon the type of cipher alphabets involved. The latter may be the secondary alphabets resulting from the sliding of the normal sequence against its reverse, or a mixed component against the normal, and so on. The student has enough information concerning the various cryptanalytic procedures which may be applied, depending upon the circumstances, in reconstructing different types of primary components and no more need be said on this score at this point.

c. The student should, however, realize one point which has thus far not been brought specifically to his attention. Although the superimposition diagram referred to in the preceding subparagraph may be composed of many columns, there is often only a relatively small number of different cipher alphabets involved. For example, in the case of two primary components of 26 letters each there is a maximum of 26 secondary cipher alphabets. Consequently, it follows that in such a case if a superimposition diagram is composed of say 100 columns, certain of those columns must represent similar secondary alphabets. There may, and probably will be no regularity of recurrence of these repeated secondaries, for they are used in a manner directly governed by the letters composing the words of the key text or the elements composing the keying sequence.

d. But the latter statement offers an excellent clue. It is clear that the number of times a given secondary alphabet is employed in such a superimposition diagram depends upon the composition of the key text. Since in the case of a running-key system using a book as a key the key text constitutes intelligible text, it follows that the various secondary alphabets will be employed with frequencies which are directly related to the respective frequencies of occurrence of letters in normal plain text. Thus, the alphabet corresponding to keyletter E should be the most frequently used; the alphabet

corresponding to keyletter T should be next in frequency, and so on. From this it follows that instead of being confronted with a problem involving as many different secondary cipher alphabets as there are columns in the superimposition diagram, the cryptanalyst will usually have not over 26 such alphabets to deal with; and allowing for the extremely improbable repetitive use of alphabets corresponding to keyletters J, K, Q, X, and Z, it is likely that the cryptanalyst will have to handle only about 19 or 20 secondary alphabets.

e. Moreover, since the E secondary alphabet will be used most frequently and so on, it is possible for the cryptanalyst to study the various distributions for the columns of the superimposition diagram with a view to assembling those distributions which belong to the same cipher alphabet, thus making the actual determination of values much easier in the combined distributions than would otherwise be the case.

f. However, if the keying sequence does not itself constitute intelligible text, even if it is a random sequence, the case is by no means hopeless of solution -- provided there is sufficient text within columns so that the columnar frequency distributions may afford indications enabling the cryptanalyst to amalgamate a large number of small distributions into a smaller number of larger distributions.

g. In this process of assembling or combining individual frequency distributions which belong to the same cipher alphabet, recourse may be had to a procedure merely alluded to in connection with previous problems, and designated as that of "matching" distributions. The next few paragraphs will deal with this important subject.

46. The "Cross-product" or "X-test" in cryptanalysis. - a. The student has already been confronted with cases in which it was necessary or desirable to reduce a large number of frequency distributions to a smaller number by

identifying and amalgamating distributions which belong to the same cipher alphabet. Thus, for example, in a case in which there are, say, 15 distributions but only, say, 5 separate cipher alphabets, the difficulty in solving a message can be reduced to a considerable degree provided that in the 15 distributions those which belong together can be identified and allocated to the respective cipher alphabets to which they apply.

b. This process of identifying distributions which belong to the same cipher alphabet involves a careful examination and comparison of the various members of the entire set of distributions to ascertain which of them present sufficiently similar characteristics to warrant their being combined into a single distribution applicable to one of the cipher alphabets involved in the problem. Now when the individual distributions are fairly large, say containing over 50 or 60 letters, the matter is relatively easy for the experienced cryptanalyst and can be made by the eye; but when the distributions are small, each containing a rather small number of letters, ocular comparison and identification of two or more distributions as belonging to the same alphabet become quite difficult and often inconclusive. In any event, the time required for the successful reduction of a multiplicity of individual small distributions to a few larger distributions is, in such cases, a very material factor in determining whether the solution will be accomplished in time to be of actual value or merely of historical interest.

c. However, a certain statistical test, called the "cross-product"/^{sum} or "X-test", has been devised, which can be brought to bear upon this question and, by methods of mathematical comparison, eliminate to a large degree the uncertainties of the ocular method of matching and combining frequency distributions, thus in many cases materially reducing the time required for solution of a complex problem.

d. It is advisable to point out, however, that the student must not expect too much of a mathematical method of comparing distributions, because there are limits to the size of distributions to be matched, below which these methods will not be effective. If two distributions contain some similar characteristics the mathematical method will merely afford a quantitative measure of the degree of similarity. Two distributions may actually pertain to the same cipher alphabet but, as occasionally happens, they may not present any external evidences of this relationship, in which case no mathematical method can indicate the fact that the two distributions are really similar and belong to the same alphabet.

47. Derivation of the χ (Chi)-test. a. Consider the following plain-text distribution of 50 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

In a previous text¹ it was shown that the chance of drawing two identical letters in normal English telegraphic plain text is the sum of the squares of the relative probabilities of occurrence of the 26 letters in such text, which is .0667. That is, the probability of monographic coincidence in English telegraphic plain text is $K_p = .0667$. In the message to which the foregoing distribution of 50 letters applies, the number of possible pairings (comparisons) that can be made between single letters is $\frac{50 \times 49}{2} = 1225$. According to the theory of coincidences there should, therefore, be $1225 \times .0667 = 81.7065$ or approximately 82 coincidences of single letters. Examining the distribution it is found that there are 83 coincidences, as shown below:

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
 $3+0+0+1+21+0+0+1+3+0+0+0+1+10+15+0+0+1+10+15+1+0+1+0+0+0 = 83$

¹ Military Cryptanalysis, Part II, Appendix 2.

- 128 -

The actual number of coincidences agrees very closely with the theoretical number, which is of course to be expected, since the text to which the distribution applied has been indicated as being normal plain text.

b. In the foregoing simple demonstration, let the number of comparisons that can be made in the distribution be indicated symbolically by $\frac{N(N-1)}{2}$, where N = the total number of letters in the distribution. Then the expected number of coincidences may be written as $\frac{.0667N(N-1)}{2}$, which may then be rewritten as

$$(I) \frac{.0667N^2 - .0667N}{2}$$

c. Likewise, if f_A represents the number of occurrences of A in the foregoing distribution, then the number of coincidences for the letter A may be indicated symbolically by $\frac{f_A(f_A-1)}{2}$. And similarly, the number of coincidences for the letter B may be indicated by $\frac{f_B(f_B-1)}{2}$, and so on down to $\frac{f_Z(f_Z-1)}{2}$.

The total number of actual coincidences found in the distribution is, of course, the sum of $\frac{f_A(f_A-1)}{2} + \frac{f_B(f_B-1)}{2} + \dots + \frac{f_Z(f_Z-1)}{2}$. If the symbol f_θ is used to indicate any of the letters A, B, ...Z, and the symbol \sum is used to indicate that the sum of all the elements that follow this sign is to be found, then the sum of the actual coincidences noted in the distribution may be indicated thus: $\sum \frac{f_\theta(f_\theta-1)}{2}$, which may be re-written as

$$(II) \sum \frac{f_\theta^2 - f_\theta}{2}$$

d. Now although derived from different sources, the two expressions labeled (I) and (II) above are equal, or should be equal, in normal plain text. Therefore, one may write:

$$\sum \frac{f_\theta^2 - f_\theta}{2} = \frac{.0667N^2 - .0667N}{2}$$

Simplifying this equation:

$$(III) \sum f_{\theta}^2 - \sum f_{\theta} = .0667N^2 - .0667N$$

e. Now $\sum f_{\theta} = N$

Therefore, expression (III) may be written as

$$(IV) \sum f_{\theta}^2 - N = .0667N^2 - .0667N,$$

which on reduction becomes:

$$(V) \sum f_{\theta}^2 = .0667N^2 + .9333N$$

This equation may be read as "the sum of the squares of the absolute frequencies of a distribution is equal to .0667 times the square of the total number of letters in the distribution, plus .9333 times the total number of letters in the distribution." The letter S_2 is often used to replace the symbol

$$\sum f_{\theta}^2.$$

f. Suppose two monoalphabetic distributions are thought to pertain to the same cipher alphabet. Now if they actually do belong to the same alphabet, and if they are correctly² combined into a single distribution, the latter must still be monoalphabetic in character. That is, again representing the individual letter frequencies in one of these distributions by the general symbol f_1 , the individual letter frequencies in the other distribution by f_{θ_2} , and the total frequency in the first distribution by N_1 , that in the second distribution by N_2 , then

$$(VI) \sum (f_{\theta_1} + f_{\theta_2})^2 = .0667(N_1 + N_2)^2 + .9333(N_1 + N_2)$$

Expanding the terms of this equation:

$$(VII) \sum f_{\theta_1}^2 + 2 \sum f_{\theta_1} f_{\theta_2} + \sum f_{\theta_2}^2 = .0667(N_1^2 + 2N_1N_2 + N_2^2) + .9333N_1 + .9333N_2$$

But from equation (V)

$$\sum f_{\theta_1}^2 = .0667N_1^2 + .9333N_1 \text{ and}$$

²By "correctly" is meant that the two distributions are slid relative to each other to their proper superimposition.

$$\sum f_{\theta_2}^2 = .0667N_2^2 + .9333N_2, \text{ so that}$$

equation (VII) may be rewritten thus:

$$.0667N_1^2 + .9333N_1 + 2 \sum f_{\theta_1} f_{\theta_2} + .0667N_2^2 + .9333N_2 =$$

$$.0667(N_1^2 + 2N_1N_2 + N_2^2) + .9333N_1 + .9333N_2$$

Reducing to simplest terms by cancelling out similar expressions:

$$2 \sum f_{\theta_1} f_{\theta_2} = .0667(2N_1N_2), \text{ or}$$

$$(VIII) \frac{\sum f_{\theta_1} f_{\theta_2}}{N_1N_2} = .0667$$

g. The last equation thus permits of establishing an expected value for the sum of the products of the corresponding frequencies of the two distributions being considered for amalgamation. The cross-product/sum or X-test for matching two distributions is based upon equation (VIII).

48. Applying the X-test in matching distributions. - a. Suppose the following two distributions are to be matched:

$f_1 \dots \overline{A} \overline{B} \overline{C} \overline{D} \overline{E} \overline{F} \overline{G} \overline{H} \overline{I} \overline{J} \overline{K} \overline{L} \overline{M} \overline{N} \overline{O} \overline{P} \overline{Q} \overline{R} \overline{S} \overline{T} \overline{U} \overline{V} \overline{W} \overline{X} \overline{Y} \overline{Z}$

$f_2 \dots \overline{A} \overline{B} \overline{C} \overline{D} \overline{E} \overline{F} \overline{G} \overline{H} \overline{I} \overline{J} \overline{K} \overline{L} \overline{M} \overline{N} \overline{O} \overline{P} \overline{Q} \overline{R} \overline{S} \overline{T} \overline{U} \overline{V} \overline{W} \overline{X} \overline{Y} \overline{Z}$

Let the frequencies be juxtaposed, for convenience in finding the sum of the cross products. Thus:

f_{θ_1}	...	1	4	0	3	0	1	0	0	1	0	0	1	0	0	1	0	0	3	2	2	1	0	1	3	0	2	...	$N_1 = 26$
		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		
f_{θ_2}	...	0	2	0	0	0	3	0	0	1	0	1	0	0	1	1	0	0	3	1	1	0	0	0	0	1	2	...	$N_2 = 17$
$f_{\theta_1} f_{\theta_2}$...	0	8	0	0	0	3	0	0	1	0	0	0	0	0	1	0	0	9	2	2	0	0	0	0	0	4		

In this case $\sum f_{\theta_1} f_{\theta_2} = 8+3+1+1+9+2+2+4 = 30$

$$N_1N_2 = 26 \times 17 = 442$$

$$\frac{\sum f_{\theta_1} f_{\theta_2}}{N_1 N_2} = \frac{30}{422} = .0711$$

b. The fact that the quotient (.0711) agrees very closely with the expected value (.0667) means that the two distributions very probably belong together or are properly matched. Note the qualifying phrase "very probably." It implies that there is no certainty about this business of matching distributions by mathematical methods. The mathematics serve only as measuring devices, so to speak, which can be employed to measure the degree of similarity that exists.

c. There are other mathematical or statistical tests for matching, in addition to the χ -test. Moreover, it is possible to go further with the χ -test and find a measure of the reliance that may be placed upon the value obtained; but these points will be left for future discussion in subsequent texts.

d. One more point will, however, here be added in connection with the χ -test. Suppose the very same two distributions in subpar. a are again juxtaposed, but with f_{θ_2} shifted one interval to the left of the position shown in the subpar. of reference. Thus:

$$f_{\theta_1} \dots \begin{cases} 1 & 4 & 0 & 3 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 3 & 2 & 2 & 1 & 0 & 1 & 3 & 0 & 2 & \dots & N_1 = 26 \\ \{ & A & B & C & D & E & F & G & H & I & J & K & L & M & N & O & P & Q & R & S & T & U & V & W & X & Y & Z & \dots \\ & B & C & D & E & F & G & H & I & J & K & L & M & N & O & P & Q & R & S & T & U & V & W & X & Y & Z & A & \dots \end{cases}$$

$$f_{\theta_2} \dots \{ 2 \ 0 \ 0 \ 0 \ 3 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 3 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 2 \ 0 \ \dots \ N_2 = 17$$

$$\text{Here } \sum f_{\theta_1} f_{\theta_2} = 2+3+2+3 = 10 \text{ and } \frac{\sum f_{\theta_1} f_{\theta_2}}{N_1 N_2} = \frac{10}{442} = .0226$$

The observed ratio (.0226) is so much smaller than the expected (.0667) that it can be said that if the two distributions pertain to the same primary components they are not properly juxtaposed. In other words, the χ -test may also be applied in cases where two or more frequency distributions must be shifted relatively in order to find their correct juxtaposition. The theory underlying

this application of the χ -test is, of course, the same as before: two mono-alphabetic distributions when properly combined will yield a single distribution which should still be monoalphabetic in character. In applying the χ -test in such cases it may be necessary to shift two 26-element distributions to various juxtapositions, make the χ -test for each juxtaposition, and take as correct that one which yields the best value for the test.

e. The nature of the problem will, of course, determine whether the frequency distributions which are to be matched should be compared (1) by direct juxtaposition, that is, setting the A to Z tallies of one distribution directly opposite the corresponding tallies of the other distribution, as in subpar. a, or (2) by shifted juxtaposition, that is, keeping the A to Z tallies of the first distribution fixed and sliding the whole sequence of tallies of the second distribution to various juxtapositions against the first.

48. Study of a situation in which the χ -test may be applied. - a. A simple demonstration of how the χ -test is applied in matching frequency distributions may now be set before the student. The problem involved is the solution of **cryptograms** enciphered according to the progressive-alphabet system (Par. 36b), with secondary alphabets derived from the interaction of two identical mixed primary components. It will be assumed that the enemy has been using a system of this kind and that the primary components are changed daily.

b. Before attacking an actual problem of this type, suppose a few minutes be devoted to a general analysis of its elements. Consider a cipher square such as that shown in Fig. 25, which is applicable to the type of problem under study. It has been arranged in the form of a deciphering square. In this square, the horizontal sequences are all identical but merely shifted relatively; the letters inside the square are plain-text letters.

ALPHABET NO.

CIPHER LETTER

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

(Plain-text letters are within the square proper.)

Figure 25.

c. If, for mere purposes of demonstration, instead of letters within the cells of the square there are placed tallies corresponding in number with the normal frequencies of the letters occupying the respective cells, the cipher square becomes as follows (showing only the 1st three rows of the square):

ALPHABET NO.

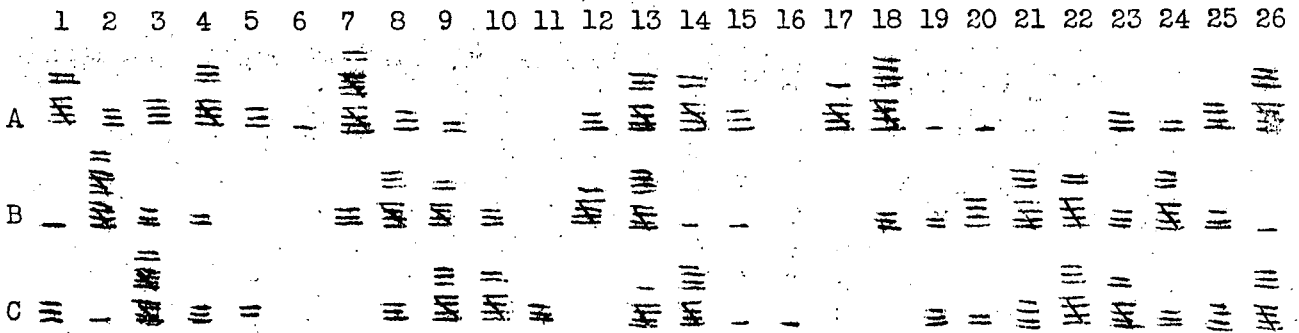


Figure 26.

d. It is obvious that here is a case wherein if two distributions pertaining to the square are isolated from the square, the X-test (matching distributions) can be applied to ascertain how the distributions can be superimposed and made to yield a monoalphabetic composite. There is obviously one correct superimposition out of 25 possibilities. In this case, the B row of tallies must be displaced 5 intervals to the right in order to match it and amalgamate it with the A row of tallies. Thus:

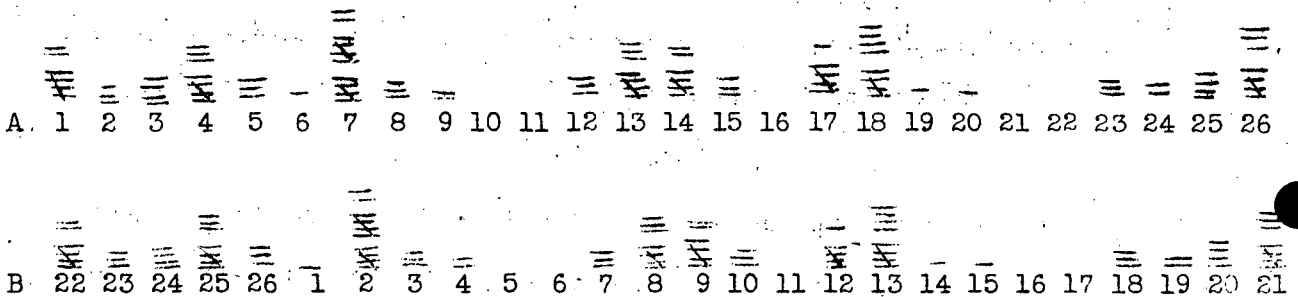


Figure 27.

e. Note that the amount of displacement, that is, the number of intervals the B sequence must be shifted to make it match the A sequence in Fig. 27, corresponds exactly to the distance between the letters A and B in the primary components, which is 5 intervals. Thus: ... A U L I C B ...

f. Given a message in such a system, suppose the cipher letters are numbered from 1 to 26, or suppose the text is transcribed in rows of 26 letters. A square table of 26 frequency distributions is then drawn up, this sort of

- 135 -

distribution merely indicating by a tally in the appropriate column whenever the cipher letter at the left appears in the text. For example, in the illustrative message above, there are no A's; the 29th letter is a B, which belongs in the 3rd alphabet; so that in the distribution square a tally is inserted in row B, column 3. The 27th letter of the message is a C, which belongs in the 1st alphabet; a tally is inserted in row C, column 1. The 37th, 38th, 39th and 53rd letters are D's; they belong in the 1st, 11th, 12th and 13th alphabets; tallies are inserted in row D, columns 1, 11, 12, and 13. Thus:

	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			1																							
C	1																									
D	1										1	1	1													

Figure 28.

The process is continued in this manner until all the letters have been distributed and represented by tallies in the appropriate rows and cells in the rows. Such a distribution may be termed a frequency distribution square.

g. When the distribution has been completed for the entire cryptogram, then the various horizontal sequences of tallies within the distribution square may be matched by the X-test and the primary components may be reconstructed, perhaps without making any assumptions of values for cipher letters. The underlying principle in this method of reconstructing primary components is that the sequences within the cipher square are identical and merely displaced various intervals, as pointed out in subparagraph c.

49. Example of solution by foregoing principles. - a. The following cryptogram has been enciphered according to the method indicated, by progressive, uninterrupted shifting of the cipher component against an identical plain component.

WGJJM	MMJXE	DGCOC	FTRPB	MIIIK	ZRYNN
BUFRW	WWWYO	IHFJK	OKHTT	AZCLJ	EPPFR
WCKOO	FFFGE	PQRY Y	IWXMX	UDIPF	EXMLL
WFKGY	PBBXC	HBFYI	ETXHF	BIVDI	PNXIV
RPWTM	GIMPT	ECJBO	KVBUQ	GVGFF	FKLYY
CKBIW	XXMUD	IPFFU	YNVSS	IHRMH	YZHAU
QWGKT	IUXYJ	JAOWZ	OCFTR	PPOQU	SGYCX
VCXUC	JLMLL	YEKFF	ZVQJQ	SIYSP	DSBBJ
UAHYN	WLOCK	SDQVC	YVSIL	IWNJO	OMAQ S
LWYJG	TVPQK	PKTLH	SROON	ICFEV	MNVWN
BNEHA	MRCRO	VSTXE	NHPVB	TWKUQ	IOCAV
WBRQN	FJVNR	VDOPU	QRLKQ	NFFFZ	PHURV
WLXGS	HQWHP	JBCNN	JQSOQ	ORCBM	RRAON
RKWUH	YYCIW	DGSJC	TGPGR	MIQMP	SGCTN
MFGJX	EDGCO	PTGPW	QQVQI	WXTTT	COJVA
AABWM	XIHOW	HDEQU	AINFK	FWHPJ	AHZIT
WZKFE	XSRUY	QIOVR	ERDJV	DKHIR	QWEDG
EBYBM	LABJV	TGFFG	XYIVG	RJYEK	FBEPB
JOUAH	CUGZL	XIAJK	WDVTY	BFRUC	CCUZZ
INNDF	RJFMB	HQLXH	MHQYY	YMWQV	CLIPT
WTJYQ	BYRLI	TUOUS	RCDCV	WDGIG	GUBHJ
VVPWA	BUJKN	FPFYW	VQZQF	LHTWJ	PDRXZ
OWUSS	GAMHN	CWHSW	WLRYQ	QUSZV	DNXAN
VNKHF	UCVVS	SSPLQ	UPCVV	VWDGS	JOGTC
HDEVQ	SIJPH	QJAWF	RIZDW	XXHCX	YCTMG
USES N	DSBBK	RLVWR	VZEEP	PPATO	IANEE
EEJNR	CZBTB	LXPJJ	KAPPM	JEGIK	RTGFF

- 137 -

HPVVV	YKJEF	HQSXJ	QDYVZ	GRRHZ	QLYXK
XAZOW	RRXYK	YGMGZ	BYNVH	QBRVF	EFQLL
WZEYL	JEROQ	SOQKO	MWIOG	MBKFF	LXDXT
LWILP	QSEDY	IOEMO	IBJML	NNSYK	XJZJM
LCZBM	SDJWQ	XTJVL	FIRNR	XHYBD	BJUFI
RJICT	UUUSK	KWDVM	FWTTJ	KCKCG	CVSAG
QBCJM	EBYNV	SSJKS	DCBDY	FPPVF	DWZMT
BPVTT	CGBVT	ZKHQD	DRMEZ	OO	

b. The message, for convenient handling, is transcribed in lines of 26 letters, as shown below in Fig. 29.

c. Frequency distributions for the 26 columns of cipher letters are made, but in a slightly different arrangement than is usually the case. Here the cryptanalyst is interested not so much in what letters occur within each cipher alphabet as he is in the occurrence of a given cipher letter in the various alphabets. The tabulation is therefore made in the form of a distribution square in which the cipher letters are shown at the left, the column or alphabet numbers at the top. Whenever a given cipher letter occurs a tally is inserted in the row marked by that letter, and in the cell in that row corresponding to the column number in which the letter occurs.

d. Since the χ -test is a purely statistical test and becomes increasingly reliable as the size of the distributions increases, it is best to start by matching the two distributions having the greatest numbers of tallies. These are the V and W distributions, with 53 and 52 occurrences, respectively. Here are the results of several trials:

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

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

Figure 29.

	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	-				-			☰	-	☰	-						☰		☰				☰	☰	-		25
B			☰	☰			☰	☰	-	-	-	-	-	-	-	-	-	-	☰	☰	☰	☰	☰	☰			43
C	☰	☰		☰	-	☰	-	-	-	-	-	-	-	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	-	45
D	-	☰	☰	☰					☰		☰			-	-	-	☰	☰	☰	☰	☰	☰	☰	☰	☰	-	34
E	☰	☰	☰	-				☰		☰				☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	35
F	☰		☰	☰		☰	☰	-	-	☰	-			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	51
G	☰	☰			-							☰	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	39
H	☰	☰			☰	☰			☰				-		☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	38
I	☰			☰	☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	45
J		-	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	50
K		☰				☰	☰	☰	☰	☰	☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	37
L	☰	☰		-	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	33
M	☰	-		☰	☰	☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	37
N	☰		☰	☰					☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	34
O	☰	☰	☰		☰	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	38
P	☰			☰	☰	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	43
Q		☰	☰	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	45
R	☰	☰	-	-	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	46
S	-		☰				☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	39
T	☰	☰	☰	-		☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	39
U	-		☰	☰	☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	33
V	-		☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	53
W	-	-	☰	☰			☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	52
X		☰		-	☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	37
Y		-			☰	☰		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	44
Z			☰	-				☰	-		☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	☰	27
	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	1042

Figure 30.

- 140 -

First Test

f_V	{	1 0 2 0 0 2 6 4 8 0 0 7 0 0 2 1 1 1 1 1 0 6 4 0 2 4	$N_V = 53$
		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	
	{	24 25 26 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	
f_W		0 4 2 1 1 5 3 0 1 0 0 2 8 1 7 6 0 1 0 0 2 3 0 2 1 2	$N_W = 52$
$f_V f_W$		0 0 4 0 0 10 18 0 8 0 0 14 0 0 14 6 0 1 0 0 0 18 0 0 2 8	$\sum f_V f_W = 103$

$$\frac{\sum f_V f_W}{N_V N_W} = \frac{103}{2756} = .037$$

Second Test

f_V	{	1 0 2 0 0 2 6 4 8 0 0 7 0 0 2 1 1 1 1 1 0 6 4 0 2 4	$N_V = 53$
		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	
	{	18 19 20 21 22 23 24 25 26 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	
f_W		2 3 0 2 1 2 0 4 2 1 1 5 3 0 1 0 0 2 8 1 7 6 0 1 0 0	$N_W = 52$
$f_V f_W$		2 0 0 0 0 4 0 16 16 0 0 35 0 0 2 0 0 2 8 1 0 36 0 0 0 0	$\sum f_V f_W = 122$

$$\frac{\sum f_V f_W}{N_V N_W} = \frac{122}{2756} = .044$$

Third Test

f_V	{	1 0 2 0 0 2 6 4 8 0 0 7 0 0 2 1 1 1 1 1 0 6 4 0 2 4	$N_V = 53$
		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	
	{	4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 1 2 3	
f_W		3 0 1 0 0 2 8 1 7 6 0 1 0 0 2 3 0 2 1 2 0 4 2 1 1 5	$N_W = 52$
$f_V f_W$		3 0 2 0 0 4 48 456 0 0 7 0 0 4 3 0 2 1 2 0 24 8 0 220	$\sum f_V f_W = 190$

$$\frac{\sum f_V f_W}{N_V N_W} = \frac{190}{2756} = .069$$

e. Since the last of the three foregoing tests gives a value somewhat better than the expected .0667, it looks as though the correct position of the W distribution with reference to the V distribution has been found. In practice, several more tests would be made to insure that other close approximations to .0667 will not be found, but these will here be omitted. The test indicates that the primary cipher component has the letters V and W in these positions:

1 2 3 4

V . . W, since the correct superimposition requires that the 4th cell of the W distribution must be placed under the 1st cell of the V distribution (see the last superimposition above).

f. The next best distribution with which to proceed is the F, with 51 occurrences. Paralleling the procedure outlined in Par. 43, and for the same reasons, the F sequence is matched against the W and V sequences separately and then against both W and V sequences at their correct superimposition. The following shows the correct relative positions of the three distributions:

f_V { 1 0 2 0 0 2 6 4 8 0 0 7 0 0 2 1 1 1 1 1 0 6 4 0 2 4 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	$N_V = 53$
f_F { 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 1 2 3 4 5 6 7 1 1 2 1 0 0 6 3 9 3 0 2 0 0 0 2 1 1 1 2 0 4 2 0 3 7	$N_F = 51$
$f_V f_F$ 1 0 4 0 0 0 3 6 1 2 2 0 0 4 0 0 0 2 1 1 1 2 0 4 8 0 6 2 8	$\sum f_V f_F = 212$

$$\frac{\sum f_V f_F}{N_V N_F} = \frac{212}{2703} = .078$$

f_W { 1 1 5 3 0 1 0 0 2 8 1 7 6 0 1 0 0 2 3 0 2 1 2 0 4 2 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	$N_W = 52$
f_F { 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 1 2 3 4 0 3 7 1 1 2 1 0 0 6 3 9 3 0 2 0 0 0 2 1 1 1 2 0 4 2	$N_F = 51$
$f_W f_F$ 0 3 3 5 3 0 2 0 0 0 4 8 3 6 3 1 8 0 2 0 0 0 6 0 2 1 4 0 1 6 4	$\sum f_W f_F = 210$

$$\frac{\sum f_W f_F}{N_W N_F} = \frac{210}{2703} = .078$$

f_{V+W} { 4 0 3 0 0 4 4 5 15 6 0 8 0 0 4 4 1 3 2 3 0 10 6 1 3 9 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	$N_{V+W} = 105$
f_F { 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 1 2 3 4 5 6 7 1 1 2 1 0 0 6 3 9 3 0 2 0 0 0 2 1 1 1 2 0 4 2 0 3 7	$N_F = 51$
$f_{V+W} f_F$ 4 0 6 0 0 0 3 4 15 13 18 0 16 0 0 0 8 1 3 2 6 0 4 12 0 9 6 3	$\sum f_{V+W} f_F = 422$

$$\frac{\sum f_{V+W} f_F}{N_{V+W} N_F} = \frac{422}{5355} = .079$$

- 142 -

The test yields the sequence $\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ V & . & . & W & . & . & . & F \end{matrix}$.

g. The process is continued in the foregoing manner until the entire primary component has been reconstructed. It is obvious that as the work progresses the cryptanalyst is forced to employ smaller and smaller distributions, so that statistically the results are apt to become less and less certain. But to counterbalance this there is the fact that the number of possible superimpositions becomes progressively smaller as the work progresses. For example, at the commencement of operations the number of possible points for superimposing a second sequence against the first is 25; after the relative positions of 5 distributions have been ascertained and a 6th distribution is to be placed in the primary sequence being reconstructed, there are 21 possible positions; after the relative positions of 20 distributions have been ascertained, there are only 6 possible positions for the 21st distribution, and so on.

h. In the foregoing case the reconstructed primary component 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
V A L W N O X F B P Y R C Q Z I G S E H T D J U M K

i. Of course, it is probable that in practical work the process of matching distributions would be interrupted soon after the positions of only a few letters in the primary components had been ascertained. For by trying partially reconstructed sequences on the cipher text the skeletons of some words would begin to show. By filling in these skeletons with the words suggested by them, the process of reconstructing the components is much facilitated and hastened.

j. The components having been reconstructed, only a moment or two is necessary to ascertain their initial position in enciphering the message. The plain text may then be obtained without further delay. It is as follows:

- 143 -

WITH THE IMPROVEMENTS IN THE AIRPLANE AND THE MEANS OF COMMUNICATION AND WITH THE VAST SIZE OF MODERN ARMIES STRATEGIC SURPRISE WILL BECOME HARDER AND HARDER TO ATTAIN X IN THE PRESENCE OF MODERN AVIATION AND FAST MOVING MECHANIZED ELEMENTS GREATER COMPLEXITIES MORE SUBTLE DECEPTIONS STRATEGEMS AND FEINTS WILL HAVE TO BE EMPLOYED X IN MODERN WARFARE IT IS STILL POSSIBLE TO GAIN TACTICAL SURPRISE BY MANY MEANS X WHILE THE MEANS OF OBSERVING AND TRANSMITTING INFORMATION OF TROOP MOVEMENTS ARE GREATLY IMPROVED OVER THOSE OF THE PAST THE MECHANICAL MEANS OF MOVING TROOPS ARE LIKEWISE FAR SPEEDIER X ALSO FALSE INFORMATION CAN BE FAR MORE EASILY AND QUICKLY DISTRIBUTED X THE LESSON TO BE LEARNED FROM THE OPENING PHASE OF ALLENBYS BATTLE OF MEGGIDO IS THAT SURPRISE IS POSSIBLE EVEN IN MODERN WARFARE BUT ONLY BY PERFECT DISCIPLINE ON THE PART OF THE TROOPS AND ALMOST SUPERHUMAN FORETHOUGHT AND ATTENTION TO DETAIL ON THE PART OF THE STAFF BACKED UP BY RESOLUTE ACTION IN THE AIR X TO MAINTAIN SECRECY MOVEMENTS MUST BE UNDER COVER OF DARKNESS AND COVERED BIVOUAC AREAS MUST BE OCCUPIED DURING DAYLIGHT HOURS X UNOBSERVED DAYLIGHT MOVEMENTS WILL REQUIRE THE RESTRICTION OF HOSTILE AIR OBSERVATION BY ANTI-AIRCRAFT ARTILLERY AND COMBAT AVIATION.

k. The student should clearly understand the real nature of the matching process employed to such good advantage in this problem. In practically all the previous cases frequency distributions were made of cipher letters occurring in a cryptogram, and the tallies in those distributions represented the actual occurrences of cipher letters. Furthermore, when these distributions were compared or matched, what were being compared were actually cipher alphabets. That is, the text was arranged in a certain way, so that letters belonging to the same cipher alphabet actually fell within the same column and the frequency distribution for a specific cipher alphabet was made by tabulating the letters in that column. Then if any distributions were to be compared, usually the entire distribution applicable to one cipher alphabet was compared with the

- 144 -

entire distribution applying to another cipher alphabet. But in the problem just completed, what were compared in reality were not frequency distributions applying to the columns of the cipher text as transcribed in Fig. 29, but were graphic representations of the variation in the frequency of use of plain-text letters falling in identical sequences, the identities of these plain-text letters being unknown for the moment. Only after the reconstruction has been completed do their identities become known, when the plain text of the cryptogram is established.

1. One final remark may be added in connection with the problem just completed. In some cases the procedure indicated, involving the matching of distributions, may not be feasible because of the paucity of text. Yet there are sufficient data in the respective alphabets to permit of some assumptions of values of cipher letters, especially if there are a few repetitions to assist in making correct assumptions. Then if the tentative values thus obtained are inserted within a reconstruction diagram applicable to the entire cipher square, it is likely that not only will the cryptanalyst be likely to find data for use in reconstructing the primary components by indirect symmetry, but also he will probably find certain data of direct assistance in the reconstruction. For example, suppose for the sake of simple illustration the cryptanalyst has been fortunate in making a good guess and has the phrase AND THE located in the foregoing message, ^{line 2,} in columns 7-12, incl. The cipher equivalents are F R W W W W. The four sequent W's mean that the primary component contains the sequence DTHE if the cipher component is sliding toward the right, or EHTD, if toward the left. (Whichever direction is correct is of no consequence, so long as the cryptanalyst is consistent in setting down the results.) Reference to the primary component reconstructed in subpar. h shows that EHTD indeed forms a sequence in that component. In other words, identical cipher

- 145 -

letters in sequence in this system mean that the plain-text letters they represent are sequent in the primary component. Going one step further, the interval between identical cipher letters in the cipher text corresponds to the interval between their equivalent plain-text letters in the primary component. Thus, the 2d and 12th cipher letters in the illustrative message above are G's; they represent I_p and O_p , respectively. Therefore, in the primary component the letters O and I are separated by 10 intervals (or 16, depending upon the direction of displacement). By applying these principles, with perhaps the addition of data from indirect symmetry, the entire primary component can often be reconstructed from a relatively small number of correct assumptions for the values of cipher letters.

50. The Φ (Phi)-test. - a. The student has noted that the X-test is based upon the general theory of coincidences and employs the probability constants K_p and K_r . There is one more test of a related nature which may be useful for him to understand and its explanation will be given in the succeeding paragraphs.

b. In Par. 47d it was stated that two monoalphabetic distributions when correctly combined will yield a single distribution which should still be monoalphabetic in character. This question arises, therefore, in the student's mind: is there a test whereby he can ascertain mathematically whether a distribution is monoalphabetic or not, especially in the case of one which has relatively few data? Such a test has been devised and is termed "the Φ (Phi) test".

51. Derivation of the Φ -test. - a. Consider a monographic or uniliteral frequency distribution which is monoalphabetic in composition. If there is a total of N letters in the distribution, in a system in which there are n possible elements, then there is a possible total of $\frac{N(N-1)}{2}$ pairs of letters (for comparison purposes).

b. Let the symbol f_A represent the number of occurrences of A, f_B the number of occurrences of B, and so on to f_Z . With regard to the letter A then, there are $\frac{f_A(f_A-1)}{2}$ coincidences. (Again the combinations of f_A things taken two at a time.) With regard to the letter B, there are $\frac{f_B(f_B-1)}{2}$ coincidences, and so on up to $\frac{f_Z(f_Z-1)}{2}$ coincidences for the letter Z. Now it has been seen that according to the K-test, in $\frac{N(N-1)}{2}$ comparisons of letters forming the two members of pairs of letters in normal English plain text, there should be $K_p \frac{N(N-1)}{2}$ coincidences, where K_p is the probability of monographic coincidence for the language in question.

c. Now the expected value of $\frac{f_A(f_A-1)}{2} + \frac{f_B(f_B-1)}{2} + \dots + \frac{f_Z(f_Z-1)}{2}$ is equal to the theoretical number of coincidences to be expected in $\frac{N(N-1)}{2}$ comparisons of two letters, which for normal plain text is K_p times $\frac{N(N-1)}{2}$ and for random text is K_r times $\frac{N(N-1)}{2}$. That is, for plain text:

$$\text{Expected value of } \frac{f_A(f_A-1)}{2} + \frac{f_B(f_B-1)}{2} + \dots + \frac{f_Z(f_Z-1)}{2} = K_p \times \frac{N(N-1)}{2}, \text{ or}$$

$$(IX) \quad \text{Expected value of } f_A(f_A-1) + f_B(f_B-1) + \dots + f_Z(f_Z-1) = K_p N(N-1);$$

and for random text:

$$\text{Expected value of } \frac{f_A(f_A-1)}{2} + \frac{f_B(f_B-1)}{2} + \dots + \frac{f_Z(f_Z-1)}{2} = K_r \times \frac{N(N-1)}{2}, \text{ or}$$

$$(X) \quad \text{Expected value of } f_A(f_A-1) + f_B(f_B-1) + \dots + f_Z(f_Z-1) = K_r N(N-1).$$

If for the left-hand side of equations (IX) and (X) the symbol $E(\Phi)$ is used, then these equations become:

$$(XI) \quad \text{For plain text ... } E(\Phi_p) = K_p N(N-1)$$

$$(XII) \quad \text{For random text .. } E(\Phi_r) = K_r N(N-1)$$

Where $E(\Phi)$ means the average or expected value of the expression in the

parenthesis, K_p and K_r are the probabilities of monographic coincidence in plain and in random text, respectively.

d. Now in normal English plain text it has been found that $K_p = .0667$. For random text of a 26-letter alphabet $K_r = .038$. Therefore, equations (XI) and (XII) may now be written thus:

$$(XIII) \quad \text{For normal English plain text } E(\phi_p) = .0667 N(N-1)$$

$$(XIV) \quad \text{For random text (26-letter alphabet) } E(\phi_r) = .0385 N(N-1)$$

e. By employing equations (XIII) and (XIV) it becomes possible, therefore, to test a piece of text for monoalphabeticity or for "randomness". That is, by using these equations one can mathematically test a very short cryptogram to ascertain whether it is a monoalphabetically enciphered substitution or involves several alphabets so that for all practical purposes it is equivalent to random text. This test has been termed the ϕ -test.

52. Applying the ϕ -test. - a. Given the following short piece of text, is it likely that it is normal English plain text enciphered monoalphabetically?

A B C D E F G H I J K L M N O P Q R S T I V W X Y Z N = 25

For this case the observed value of ϕ is:

$$(1 \times 0) + (1 \times 0) + (2 \times 1) + (3 \times 2) + (4 \times 3) + (2 \times 1) + (1 \times 0) + (4 \times 3) + (2 \times 1) + (1 \times 0) + (1 \times 0) + (3 \times 2) = 2 + 6 + 12 + 2 + 12 + 2 + 6 = 40$$

If this text were monoalphabetically enciphered English plain text the expected value of ϕ is:

$$E(\phi_p) = K_p N(N-1) = .0667 \times 25 \times 24 = 40.0$$

If the text were random text, the expected value of ϕ is:

$$E(\phi_r) = K_r N(N-1) = .0385 \times 25 \times 24 = 23.1$$

The conclusion is warranted, therefore, that the cryptogram is probably monoalphabetic substitution, since the observed value of ϕ (40) more closely approximates the expected value for English plain text (40.0) than it does the

expected value for random text (23.1). (As a matter of fact, the cryptogram was enciphered monoalphabetically.)

b. Here is another example. Given the following series of letters, does it represent a selection of English text enciphered monoalphabetically or does it more nearly represent a random selection of letters?

Y O U I J Z M M Z Z M R N Q C X I Y T W R G K L H

The distribution and calculation are as follows:

	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
f(f-1)...	0						0	0	2	0	0	0	6	0	0	0	2		0	0		0	0	2	6	

$$\sum f(f-1) = 18 \text{ (That is, observed value of } \phi = 18)$$

$$E(\phi_p) = .0667 \times 25 \times 24 = 40.0 \text{ (That is, expected value of } \phi_p = 40.0)$$

The conclusion is that the series of letters does not represent a selection of English text monoalphabetically enciphered. Whether or not it represents a random selection of letters cannot be told, but it may be said that if the letters actually do constitute a cryptogram, the latter is probably polyalphabetically enciphered. (As a matter of fact, the latter statement is true, for the message was enciphered by 25 alphabets used in sequence.)

c. The ϕ -test is, of course, closely related to the X-test and derives from the same general theory as the latter, which is that of coincidence.

When two monoalphabetic distributions have been combined into a single distribution, the ϕ -test may be applied to the latter as a check upon the X-test.

It is also useful in testing the columns of a superimposition diagram, to ascertain whether or not the columns are monoalphabetic.

SECTION XIII

CONCLUDING REMARKS

	Paragraph
Concluding remarks on aperiodic substitution systems.....	53
Synoptic table.....	54

53. Concluding remarks on aperiodic substitution systems. - a. The various systems described in the foregoing pages represent some of the more common and well-known methods of introducing complexities in the general scheme of cryptographic substitution with the view to avoiding or suppressing periodicity. There are, of course, other methods for accomplishing this purpose, which, while perhaps a bit more complex from a practical point of view, yield more desirable results from a cryptographic point of view. That is, these methods go deeper into the heart of the problem of cryptographic security and thus make the task of the enemy cryptanalyst much harder. But studies based on these more advanced methods will have to be postponed at this time, and reserved for a later text.

b. Thus far in these studies, aside from a few remarks of a very general nature, no attention has been paid to that other large and important class of ciphers, viz, transposition. It is desirable, before going further with substitution methods, that the student gain some understanding of how to solve certain of the more simple varieties of transposition ciphers. Consequently, in the text to succeed the present text, the student will temporarily lay aside the various useful methods and tools that he has been given for the solution of substitution ciphers and will turn his thoughts toward the methods of breaking down transposition ciphers.

54. Synoptic table. - Continuing the plan instituted in previous texts, of summarizing the textual material in the form of a very condensed chart called An Analytical Key for Military Cryptanalysis, the outline for the studies covered by Part III follows herewith.

INDEX

	Par.	Pages
Aperiodic systems	<u>1b</u>	1
Arithmetical equivalent of normal sliding-strip encipherment.....	29c	65
Auto-key encipherment, two basic methods of.....	23	46
Auto-key systems, solution of.....	23-38	46-88
Auto-key systems, characteristics of.....	25h-j, 38b	55-56, 88
Auto-key systems, concluding remarks on.....	38	88
Auto-keying, cipher text.....	23c	46
Auto-keying, plain text	23b, 27	46, 59
Avoiding periodicity, methods of	1-4, 17	1-8, 31
Base letter	25j	56
Basic period masked by aperiodic repetitions	16c	27
Blocking out isomorphs	14c	23
Blocking out words	14a	22
Book as key	5b	8
Chi-test, applying the	47a	130
Chi-test, derivation of	46e	127
Chi-text, example of application of	48-49	132-145
Chi-test in matching shifted distributions	47d	131
Coincidence test	41c	106
Coincidence test, application of	43	110-119
Coincidence test, basic theory of	41	101-110
Coincidence, intermittent	16e	29
Combining individual frequency distributions	45g	125
Comparisons for coincidence.....	41d	106
Constant-length, plain-text groupings	2a	2
Continuous-key system.....	36	91
Conversion of an aperiodic cipher into periodic form	32	81-88
Converting auto-key text to periodic terms	32	81-88
Cross-product or X-test	46-49	125-145
Cryptanalytic coincidence test	41-44	101-123
Cryptographic arithmetic	29c	65
Cryptographic periodicity, nature of.....	2	2
Cyclic phenomena	1a	1
Enciphering equations	9b	12
Encipherment by word lengths	6a	9
Extended keys	35	89
Formulae, idiomorphic	10c	14
Frequency distribution square	48k	135
General solution for ciphers involving a long keying sequence	40	99
Groupings, constant-length plain-text	17-22	31-45
Groupings, irregular.....	16	26
Groupings, variable-length, plain-text	5-16	8-30

	Par.	Pages
Identity or coincidence	41c	103
Idiomorphism	10b	13
Indicators	40b	100
Influence letter.....	18e(note)	34
Interlocking messages by repetitions	40b	100
Intermittent coincidences	16f	30
Interrupting a cyclic keying sequence	18	31
Interrupting the key, three basic methods of	18d	34
Interruptions, keying	17	31
Interruptor	18d	34
Interruptor, cipher text letter as	21	41
Interruptor methods, disadvantages of	22e	44
Introductory key.....	25h	55
Introductory key consisting of more than one letter	30	74-78
Introductory key, finding the length of the	30	74-78
Irregular interruptions in keying sequence	17	31
Isomorphic sequences.....	12a	18
Isomorphism, illustration of the use of	13	18
Isomorphism, phenomena of	11c,14e	18,25
Isomorphisms, detection of	14c	23
Isomorphs, blocking out of	15b	26
Kappa test	41c	106
Kappa test, application of	43a	110
Keying, fixed	2b	3
Keying cycles, interaction of	3c	5
Keying units, constant length	5-16	8-30
Keying units, variable length	17-22	31-45
Keys, extended; non-repeating; running	35	89
Lengthening keys	34	89
Lengthy keys, systems using	1c	2
Lengthy keys, mechanical methods of producing	36d	92
Making the K-test, general procedure.....	42	108
Matching of frequency distributions	45-52	123-148
Monographic coincidence, probability of	41b	101
Non-repeating key system	35-36	89-91
Overlap	36a,38b	92,97
Patterns, idiomorphic	10c	14
Patterns, word	10b	13
Period, apparent; basic; complete; hidden; latent; patent; primary; resultant; secondary	4	7
Periodicity, masked	16c	27
Periods, component.....	4a	7
Periods, superimposed	20a	38
Phi-test, applying	52	147-148
Phi-test, derivation of	51	145-147
Phi-test, nature of	50a	145-148
Phi-test related to Chi-test	52c	148

	Par.	Pages
Probability, theory of	41a	101
Probability of monographic coincidence	46e	127
Probable-word method	10b, 37b	13, 94
Progressive alphabet cipher, solution of	39	97
Progressive alphabet system	36b	92
Reconstruction skeleton	10d	14
Repetitions, completely periodic	3b	5
Repetitions, nonperiodic	16c	27
Repetitions, partially periodic	3b	5
Repetitions, significant	21b	42
Resultant key	36d	93
Running-key cipher, solution of	37-40, 44c	94-101, 120
Running-key system	36a	91
Secondary key.....	36d	93
Separators, word	15a	25
Sequences, uninterrupted	21d	42
Solution by superimposition	19f, 20b, 40a	37, 39, 99
Spurious plain text.....	29f	70
Statistical test.....	22b	43
Stereotypic phraseology	14b	22
Superimposable periods	20a	38
Superimposed sequences and the coincidence test....	41c	102
Superimposition	36c	92
Superimposition, basic principles of.....	20	37-41
Superimposition, correct and incorrect	41c	103
Superimposition, solution by	20	37-41
Superimposition diagram	41f	107
Synoptic table	54	149
Symmetry of position, direct	10e	15
Variable-length groupings of keying sequence	17	31
Variable-length key enciphering	22d	44
Variable-length plain-text groupings	3, 16d, 22d	3, 27, 44
Vigenère method	28c	61
Wheatstone cryptograph	36e	93
Word habits of the enemy, familiarity with	20b	41
Word-length encipherment, solution of	5-9	8-12
Word separators	15	25

ANALYTICAL KEY FOR MILITARY CRYPTANALYSIS, PART III
(Numbers in parentheses refer to Paragraph Numbers in this text)

506-Aperiodic Systems
(1-54)

613-Constant length keying units encipher variable length plain-text groupings (2-16).

614-Variable length keying units encipher constant-length plain-text groupings (17-22).

615-Auto key Systems (23-33).

616-Systems using lengthy keys (34-54).

711-Variable length plain-text groupings other than word lengths (2,3,16).

712-Word length plain-text groupings (6,10-13).

713-Interruptor is a plain-text letter (19-20).

714-Interruptor is a cipher-text letter (21-22).

715-Cipher text auto-keying (23-26).

716-Plain text auto-keying (27-33).

717-Running-key systems (35,36a,37-38,40-54).

718-Progressive alphabet systems (36b,39-40,41-54).

813-Original plain-text groupings are retained in the cryptograms (10-13).

814-Original plain-text groupings are not retained in the cryptograms (14-15).

815-Introductory key is a single letter (23,26,27-29).

816-Introductory key is a word or phrase (23,25h,26,30-33).

905-With standard or known cipher alphabets (6-9).

906-With mixed or unknown cipher alphabets (10-15).