For Mr Friedman

This is a revised copy
of subject paper.

SECTION IX

POLYGRAPHIC SUBSTITUTION SYSTEMS

64. General remarks on polygraphic substitution.--a. The substitution systems dealt with thus far have involved plaintext units consisting of single elements (usually single letters). The major distinction between them has been made simply on the basis of the number of elements constituting the ciphertext units of each; i.e., those involving single-element ciphertext units were termed uniliteral, and those involving ciphertext units composed of two or more elements were termed multiliteral.[1] That is to say, when the terms "uniliteral", "biliteral", "triliteral", etc., were used, it was to have been automatically inferred that the plaintext units were composed of single elements.

b. This section of the text will deal with substitution systems involving plaintext units composed of more than one element; such systems are termed polygraphic.[2] (By comparing this new term with the terms "uniliteral" and "multiliteral" it may then be deduced--and correctly so--that a term involving the suffix "-literal" is descriptive of the composition of the cipher text units of a cryptosystem, and that a term containing the suffix "-graphic" describes the composition of the

[1] See also subpar. 52a.

[2] Systems involving plaintext units composed of single elements may, on this basis, be termed monographic; however, as has been stated in connection with the terms "uniliteral" and "multiliteral", the plaintext units of a system are understood (without restatement) to be monographic unless otherwise specified.

plaintext units.[3]) Polygraphic systems in which the plaintext units are composed of two elements are called *digraphic*, those in which the plaintext units are composed of three elements are *trigraphic*, etc. The *ciphertext* units of polygraphic systems usually consist of the same number of elements as the plaintext units.[4] Thus, if a system is called "digraphic", it may be assumed that the ciphertext units of the system consist of two elements, as do the plaintext units; if this were not the case, the term "digraphic" by itself would not be adequate to describe the system completely, and an additional modifying word or phrase would have to be used to indicate this fact.[5]

c. In polygraphic substitution, the *combinations* of elements which constitute the plaintext units are considered as indivisible compounds. The units are composite in character and the individual elements composing the units affect the equivalent cipher units *jointly*, rather than separately. The basic important factor in true polygraphic substitution is that *all* the letters of each plaintext unit participate in the determination of its cipher equivalent; the identity of *each* element of the plaintext unit affects the composition of the *whole* cipher unit.[6] Thus, in a certain digraphic system, $\overline{AB}_p$ may be enciphered as $\overline{XP}_c$; and $\overline{AC}_p$, on the other hand, may be enciphered as $\overline{NK}_c$; a difference in the identity of but one of the letters of the plaintext pair here produces a difference in the identity of *both* letters of the cipher pair.[7]

[3] In this connection, it is further pointed out that since the root "literal" derives from the Latin "litera", it is conventionally prefixed by modifiers of Latin origin, such as "uni-", "bi-", and "multi-"; similarly, "graphic", deriving from the Greek "graphikos", is prefixed by modifiers of Greek origin, such as "mono-", "di-", and "poly-".

[4] The qualifying adverb "usually" is employed because this correspondence is not essential. For example, if one should draw up a set of 676 arbitrary single signs, it would be possible to represent the 2-letter pairs from AA to ZZ by single symbols. This would still be a digraphic system.

[5] See subpars. 65*e* and 66*f* for examples of two such systems and their names.

[6] An analogy is found in chemistry, when two elements combine to form a molecule, the latter usually having properties quite different from those of either of the constituent elements. For example: sodium, a metal, and chlorine, a gas, combine to form sodium chloride, common table salt. However, sodium and fluorine, also a gas similar in many respects to chlorine, combine to form sodium fluoride, which is much different from table salt.

[7] For this reason the two letters are marked by a ligature; that is, by a bar across their tops. In cryptologic notation, the symbol $\overline{\theta\theta}_p$ means "any plaintext digraph"; the symbol $\overline{\theta\theta}_c$, "any ciphertext digraph". To refer specifically to the 1st, 2d, 3d,... member of a ligature, the exponent 1, 2, 3,... will be used. Thus θ_p^2 of $\overline{REM}_p$ is the letter E; θ_c^3 of $\overline{XRZ}_c$ is Z. See also footnote 1 on page [illegible].

d. The fundamental purpose of polygraphic substitution is again the suppression or the elimination of the frequency characteristics of single letters of plain text, just as is the case in monoalphabetic substitution with variants; but here this is accomplished by a different method, the latter arising from a somewhat different approach to the problem involved in producing cryptographic security. When the substitution involves replacement of single letters in a monoalphabetic system, even a single cryptogram can be solved rather readily; basically the reason for this is that the principles of frequency and the laws of probability, applied to individual units (single letters) of the plain text, have a very good opportunity to manifest themselves. However, when the substitution involves replacement of plaintext units composed of two or more letters--that is, when the substitution is polygraphic in nature--the principles of frequency and laws of probability have a much lesser opportunity to manifest themselves. If the substitution is digraphic, then the units are pairs of letters and the normal frequencies of plaintext digraphs become of first consideration; if the substitution is trigraphic, the units are sets of three letters and the normal frequencies of plaintext trigraphs are involved. In these cases the data that can be employed in the solution are meager; that is why, generally speaking, the solution of polygraphic substitution ciphers is often extremely difficult.

e. By way of example, a given plaintext message of say n letters, enciphered by means of a uniliteral substitution system, affords n cipher characters, and the same number of cipher units. The same message, enciphered digraphically, still affords n cipher characters but only $\frac{n}{2}$ cipher units. Statistically speaking, the sample to which the laws of probability now are to be applied has been cut in half. Furthermore, from the point of view of frequency, the very noticeable diversity in the frequencies of individual letters, leading to the marked crests and troughs of the uniliteral frequency distribution, is no longer so strikingly in evidence in the frequencies of digraphs. Therefore, although digraphic encipherment, for example, simply cuts the cryptographic textual units in half, the number of cipher units which must be identified has been squared; and the difficulty of solution is not merely doubled but, if a matter of judgment arising from practical experience can be expressed or approximated mathematically, squared or cubed.

f. The following two paragraphs will treat various polygraphic substitution methods. The most practical of these methods are digraphic in character and for this reason their treatment herein will be more detailed than that of trigraphic methods.

65. Polygraphic substitution methods employing large tables.--

a. The simplest method of effecting polygraphic substitution involves the use of tables similar to that shown in Figure 47a. This table merely provides equivalents for digraphs, by means of the coordinate system. Specifically, in obtaining the cipher equivalent of any

	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	WG	EE	SN	TR	IA	NL	GC	HT	OI	UO	AM	RP	BY	KB	CD	DF	FH	JJ	LK	MQ	PS	QU	VV	XW	YX	ZZ
B	EG	SE	TN	IR	NA	GL	HC	OT	UI	AO	RM	BP	KY	CB	DD	FF	JH	LJ	MK	PQ	QS	VU	XV	YW	ZX	WZ
C	SG	TE	IN	NR	GA	HL	OC	UT	AI	RO	BM	KP	CY	DB	FD	JF	LH	MJ	PK	QQ	VS	XU	YV	ZW	WX	EZ
D	TG	IE	NN	GR	HA	OL	UC	AT	RI	BO	KM	CP	DY	FB	JD	LF	MH	PJ	QK	VQ	XS	YU	ZV	WW	EX	SZ
E	IG	NE	GN	HR	OA	UL	AC	RT	BI	KO	CM	DP	FY	JB	LD	MF	PH	QJ	VK	XQ	YS	ZU	WV	EW	SX	TZ
F	NG	GE	HN	OR	UA	AL	RC	BT	KI	CO	DM	FP	JY	LB	MD	PF	QH	VJ	XK	YQ	ZS	WU	EV	SW	TX	IZ
G	GG	HE	ON	UR	AA	RL	BC	KT	CI	DO	FM	JP	LY	MB	PD	QF	VH	XJ	YK	ZQ	WS	EU	SV	TW	IX	NZ
H	HG	OE	UN	AR	RA	BL	KC	CT	DI	FO	JM	LP	MY	PB	QD	VF	XH	YJ	ZK	WQ	ES	SU	TV	IW	NX	GZ
I	OG	UE	AN	RR	BA	KL	CC	DT	FI	JO	LM	MP	PY	QB	VD	XF	YH	ZJ	WK	EQ	SS	TU	IV	NW	GX	HZ
J	UG	AE	RN	BR	KA	CL	DC	FT	JI	LO	MM	PP	QY	VB	XD	YF	ZH	WJ	EK	SQ	TS	IU	NV	GW	HX	OZ
K	AG	RE	BN	KR	CA	DL	FC	JT	LI	MO	FM	QP	VY	XB	YD	ZF	WH	EJ	SK	TQ	IS	NU	GV	HW	OX	UZ
L	RG	BE	KN	CR	DA	FL	JC	LT	MI	PO	QM	VP	XY	YB	ZD	WF	EH	SJ	TK	IQ	NS	GU	HV	OW	UX	AZ
M	BG	KE	CN	DR	FA	JL	LC	MT	PI	QO	VM	XP	YY	ZB	WD	EF	SH	TJ	IK	NQ	GS	HU	OV	UW	AX	RZ
N	KG	CE	DN	FR	JA	LL	MC	PT	QI	VO	XM	YP	ZY	WB	ED	SF	TH	IJ	NK	GQ	HS	OU	UV	AW	RX	BZ
O	CG	DE	FN	JR	LA	ML	PC	QT	VI	XO	YM	ZP	WY	EB	SD	TF	IH	NJ	GK	HQ	OS	UU	AV	RW	BX	KZ
P	DG	FE	JN	LR	MA	PL	QC	VT	XI	YO	ZM	WP	EY	SB	TD	IF	NH	GJ	HK	OQ	US	AU	RV	BW	KX	CZ
Q	FG	JE	LN	MR	PA	QL	VC	XT	YI	ZO	WM	EP	SY	TB	ID	NF	GH	HJ	OK	UQ	AS	RU	BV	KW	CX	DZ
R	JG	LE	MN	PR	QA	VL	XC	YT	ZI	WO	EM	SP	TY	IB	ND	GF	HH	OJ	UK	AQ	RS	BU	KV	CW	DX	FZ
S	LG	ME	PN	QR	VA	XL	YC	ZT	WI	EO	SM	TP	IY	NB	GD	HF	OH	UJ	AK	RQ	BS	KU	CV	DW	FX	JZ
T	MG	PE	QN	VR	XA	YL	ZC	WT	EI	SO	TM	IP	NY	GB	HD	OF	UH	AJ	RK	BQ	KS	CU	DV	FW	JX	LZ
U	PG	QE	VN	XR	YA	ZL	WC	ET	SI	TO	IM	NP	GY	HB	OD	UF	AH	RJ	BK	KQ	CS	DU	FV	JW	LX	MZ
V	QG	VE	XN	YR	ZA	WL	EC	ST	TI	IO	NM	GP	HY	OB	UD	AF	RH	BJ	KK	CQ	DS	FU	JV	LW	MX	PZ
W	VG	XE	YN	ZR	WA	EL	SC	TT	II	NO	GM	HP	OY	UB	AD	RF	BH	KJ	CK	DQ	FS	JU	LV	MW	PX	QZ
X	XG	YE	ZN	WR	EA	SL	TC	IT	NI	GO	HM	OP	UY	AB	RD	BF	KH	CJ	DK	FQ	JS	LU	MV	PW	QX	VZ
Y	YG	ZE	WN	ER	SA	TL	IC	NT	CI	HO	OM	UP	AY	RB	BD	KF	CH	DJ	FK	JQ	LS	MU	PV	QW	VX	XZ
Z	ZG	WE	EN	SR	TA	IL	NC	GT	HI	OO	UM	AP	RY	BB	KD	CF	DH	FJ	JK	LQ	MS	PU	QV	VW	XX	YZ

Figure 47a.

plaintext digraph, the initial letter of the plaintext digraph is used to indicate the row in which the equivalent is found, and the final letter of the plaintext digraph indicates the column; the cipher digraph is then found at the intersection of the row and column thus indicated. For example, $\overline{KG}_p=\overline{FC}_c$; $\overline{WM}_p=\overline{OY}_c$; etc.

b. In the preceding table two mixed sequences were employed to form the cipher equivalents, one sequence being based on the key phrase WESTINGHOUSE AIR BRAKE and the other on GENERAL ELECTRIC COMPANY. The table in Figure 47a could have been drawn up in a slightly different manner, as shown in Figure 47b, and still yield the same cipher equivalents as before. Using this latter table, θ_c^1 for any plaintext digraph is found at the intersection of the row and column identified by θ_p^1 and θ_p^2, respectively; θ_c^2 is found in the sequence below the table and is

θ^2_p

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

Figure 47b.

taken from the position directly under the column identified by θ^2_p. A few sample encipherments will illustrate that this table is cryptographically equivalent to that of Fig. 47a.

c. Figures 48 and 49, below, contain other possible types of tables for digraphic substitution. In Fig. 48, it will be seen that there are two vertical sequences to the left of this table and no horizontal sequence below it. θ^1_p is located in the leftmost sequence, θ^1_c being found directly to its side in the right-hand sequence; θ^2_c is then found at the intersection of the row and column identified by θ^1_p and θ^2_p,

θ^2_p

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

Figure 48.

respectively. The table in Fig. 49 provides digraphic equivalents by means of the coordinate system (e.g., $\overline{RE}_p = \overline{JZ}_c$), in the same manner as in Fig. 47a, and a cursory examination of the inside of the table might disclose nothing new about this table at all. But, if one were to scan closely the diagonals formed by each θ^1_c from upper right to lower left, he would see that each such diagonal changes below the "M_p row"; similarly, if the diagonals formed by θ^2_c are scanned from upper left to lower right, it will be seen that each of them also changes after the "M_p row". In effect, the inside of the table is divided into two separate portions by an imaginary line extending horizontally between the M and N rows; but within each portion a straightforward type of symmetry is exhibited and the same two mixed sequences have been employed in each. Actually, in a 26x26 table, it is not possible to maintain the diagonals formed thus by θ^1_c and θ^2_c in a completely "unbroken" sequence without producing repeated digraphs within the table and without consequent cryptographic ambiguity; thus, Fig. 49 illustrates one type of limited diagonal symmetry which must be resorted to in the systematic construction of such a table.

θ^2_p

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

Figure 49.

d. All of the foregoing tables have exhibited a symmetry in the arrangement of their contents, which is undesirable from the standpoint of cryptographic security. This systematic internal arrangement could be detected by a cryptanalyst early in his attack on cryptograms produced through their use, permitting rapid reconstruction of the particular table involved; this subject will be given a more detailed treatment in par. 72. The table in Figure 50 is an example of one type of table which would provide more security than the foregoing. This table is constructed by random assignment of values and shows no symmetry whatsoever in its arrangement of contents. It will be noted that this table is

(Showing only a partially filled table)

Final Letter (θ^2_p)

Initial letter (θ^1_p)	A	B	C	D	E	F	G	H	I	J	K	. . .	X	Y	Z	
A	FX	CH	XE	YY	ZA	YG	FB	CD	EF	XJ	ZX	. . .	EA	DJ	FH	A
B	NY	DC	NB	ZI	XX	DX						. . .				B
C				AH				AB				. . .		ND		C
D			BB		YA					AY		. . .	BF			D
E	AX					AI						. . .				E
F		AG			NZ			AZ				. . .	AA			F
	:	:	:	:	:	:	:	:	:	:	:	. . .	:	:	:	
N		BC		CY								. . .		BA	FE	N
	:	:	:	:	:	:	:	:	:	:	:	. . .	:	:	:	
X					AC					AJ		. . .	BE			X
Y	DE						AF					. . .		AD		Y
Z	AE								BD			. . .	AK			Z
	A	B	C	D	E	F	G	H	I	J	K		X	Y	Z	

Figure 50.

reciprocal in nature; that is $\overline{AF}_p=\overline{YG}_c$ and $\overline{YG}_p=\overline{AF}_c$. Thus, this single table serves for deciphering as well as for enciphering. Reciprocity is, however, not an essential factor; in fact, greater security is provided by non-reciprocal tables. But, in the case of such non-reciprocal, randomly constructed tables, each enciphering table must have its complementary deciphering table.

e. Digraphic tables employing numerical equivalents instead of letter equivalents may be encountered. However, since 676 equivalents are required (there being 676, or 26x26, different pairs of letters), this means that combinations of three figures must be used; such systems are termed trinome-digraphic systems, indicating clearly the number of elements which comprise the cipher units. By way of an example, the following figure contains a fragment of a table[8] which provides trinome equivalents for the plaintext digraphs:

	A	B	C	D	E		Y	Z
A	001	002	003	004	005	....	025	026
B	027	028	029	030	031		051	052
C	053	054						
	...							
	...							
	...							
Y	625	626					649	650
Z	651	652					675	676

Figure 51.

f. All of the foregoing tables have been digraphic in nature, but a kind of false trigraphic substitution may also be accomplished by means of similar tables, as illustrated in Figure 52, wherein the table is the same as that in Figure 49 with the addition of one more sequence at the top of the table. In using this table, θ_p^1 is located in sequence I, and

[8] It is interesting to note that this comparatively bulky and unwieldy table can be reduced to the following two alphabets with numerical equivalents for the letters:

(1)	A	B	C	D	E	F	.	.	.	.	.	X	Y	Z
	000	026	052	078	104	130	...	...	...	...	...	598	624	650
(2)	A	B	C	D	E	F	.	.	.	.	.	X	Y	Z
	1	2	3	4	5	6	.	.	.	.	.	24	25	26

In enciphering, the first letter of the plaintext digraph is converted into its numerical value from alphabet (1), and the second plaintext letter is converted by means of alphabet (2); the two numerical values thus derived are added together, and their sum is taken as the cipher equivalent of the particular plaintext digraph. Of course, this simple reduction would not be possible if the trinomes, in ascending order, had been arranged in the table in, say, a diagonal manner.

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

Figure 52.

its equivalent, θ_c^1, taken from sequence II; θ_p^2 is located in sequence III, and its equivalent, θ_c^2, taken from sequence IV; θ_c^3 is the letter lying at the intersection of the row indicated by θ_p^3 in sequence I and the column determined by θ_p^2. Thus, FIRE LINES would be enciphered $\overline{\text{NNZ}}$ $\overline{\text{IEQ}}$ $\overline{\text{KOV}}$. Various other agreements may be made with respect to the alphabets in which each plaintext letter will be sought in such a table, but the basic cryptographic principles are the same as in the case described.

g. Tables such as those illustrated in Figs. 47-52, above, have been encountered in operational systems, but their use has not been very widespread because of their relatively large size and the inconvenience in their production and handling. In lieu of these large tables it is possible to employ much smaller matrices or geometrical designs to accomplish digraphic substitution; methods involving their use will be discussed in the following paragraph.

66. Polygraphic substitution methods employing small matrices.[9]--

a. A simple method for accomplishing digraphic substitution involves the use of the four-square matrix, a matrix consisting of four 5x5 squares in which the letters of a 25-element alphabet (combining I and J) are inserted in any prearranged order. When four such squares are arranged in a matrix as shown in Figure 53, the latter may be employed for digraphic substitution to yield the same cipher results as does a much larger table of the type treated in the preceding paragraph. In a four-square matrix, θ_p^1 of $\overline{\theta\theta}_p$ is sought in section 1; θ_p^2, in section 2. Thus, θ_p^1 and θ_p^2 will always form the northwest-southeast corners of an imaginary rectangle delimited by these two letters as located in these two sections of the square. Then θ_c^1 and θ_c^2 are, respectively, the letters at the northeast-southwest corners of this same rectangle. Thus, $\overline{TG}_p=\overline{XS}_c$; $\overline{WD}_p=\overline{CH}_c$; $\overline{OR}_p=\overline{YV}_c$; $\overline{UR}_p=\overline{XB}_c$; etc. In decrypting, θ_c^1 and θ_c^2 are sought in sections 3 and 4, respectively, and their equivalents, θ_p^1 and θ_p^2, noted in sections 1 and 2, respectively.

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

Figure 53.

b. It is possible to effect digraphic substitution with a matrix consisting of but two sections by a modification in the method of finding equivalents. In a horizontal two-square matrix, such as that shown in Figure 54, θ_p^1 of $\overline{\theta\theta}_p$ is located in the square at the left; θ_p^2, in the square at the right.

[9] The word matrix as employed in this paragraph refers to checkerboard-type diagrams smaller than the tables illustrated in the preceding paragraph. These matrices are usually composed of sections containing 25 cells each.

	M	A	N	U	F	A	U	T	O	M	
	C	T	R	I	G	B	I	L	E	S	
$\theta_p^1\theta_c^2$	B	D	E	H	K	C	D	F	G	H	$\theta_p^2\theta_c^1$
	L	O	P	Q	S	K	N	P	Q	R	
	V	W	X	Y	Z	V	W	X	Y	Z	

Figure 54.

When θ_p^1 and θ_p^2 are at the opposite ends of the diagonal of an imaginary rectangle defined by these letters, the ciphertext equivalent comprises the two letters appearing at the opposite ends of the other diagonal of the same rectangle; θ_c^1 is the particular one which is in the same row as θ_p^1, and θ_c^2 is the one in the same row as θ_p^2. For example, $\overline{AL}_p = \overline{TT}_c$; $\overline{DO}_p = \overline{GA}_c$. When θ_p^1 and θ_p^2 happen to be in the same row, the ciphertext equivalent is merely the reverse of the plaintext digraph; for example, $\overline{AT}_p = \overline{TA}_c$ and $\overline{EH}_p = \overline{HE}_c$.

c. Digraphic substitution may also be effected by means of vertical two-square matrices, in which one section is directly above the other, as in Figure 55; it will be noted that matrices of this type have a feature of reciprocity when employed according to the usual rules, which follow.

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

Figure 55.

When θ_p^1 and θ_p^2 are at the opposite ends of a diagonal, the rule for encipherment is the same as that for horizontal two-square encipherment (e.g., $\overline{MO}_p=\overline{UA}_c$ and $\overline{UA}_p=\overline{MO}_c$); when both θ_p^1 and θ_p^2 happen to be in the same column, the plaintext digraphs are self-enciphered, (e.g., $\overline{MA}_p=\overline{MA}_c$ and $\overline{EL}_p=\overline{EL}_c$), a fact which constitutes an important weakness of this method.[10] This disadvantage is only slightly less obvious in the preceding case of horizontal two-square methods wherein the cipher equivalent of $\overline{\theta\theta}_p$ consists merely of the plaintext letters in reversed order.

d. One-square digraphic methods, with a necessary modification of the method for finding equivalents, are also possible. The first of this type to appear as a practical military system was that known as the Playfair cipher.[11] It was used for a number of years as a field cipher by the British Army, before and during World War I, and for a short time, also during that war, by certain units of the American Expeditionary Forces. Figure 56 shows a typical Playfair square. The modification in the method of finding cipher equivalents has been found useful in imparting a greater degree of security than that afforded in the preceding small matrix methods. The usual method of encipherment can be best explained by examples given under four categories:

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

Figure 56.

[10] See subpar. 73b on other enciphering conventions which remove this weakness.

[11] This cipher was really invented by Sir Charles Wheatstone but receives its name from Lord Playfair, who apparently was its sponsor before the British Foreign Office. See Wemyss Reid, Memoirs of Lyon Playfair, London, 1899. It is of interest to note that, to students of electrical engineering, Wheatstone is generally not known for his contributions to cryptography but is famed for something he did not invent--the so-called Wheatstone bridge", really invented by Samuel H. Christie.

(1) Members of the plaintext pair, θ_p^1 and θ_p^2, are at opposite ends of the diagonal of an imaginary rectangle defined by the two letters; the members of the ciphertext pair, θ_c^1 and θ_c^2, are at the opposite ends of the other diagonal of this imaginary rectangle. Examples: $\overline{MO}_p=\overline{AL}_c$; $\overline{MI}_p=\overline{UC}_c$; $\overline{LU}_p=\overline{QM}_c$; $\overline{VI}_p=\overline{YC}_c$.

(2) θ_p^1 and θ_p^2 are in the same row; the letter immediately to the right of θ_p^1 forms θ_c^1; the letter immediately to the right of θ_p^2 forms θ_c^2. When either θ_p^1 or θ_p^2 is at the extreme right of the row, the first letter in the row becomes its cipher equivalent. Examples: $\overline{MA}_p=\overline{AN}_c$; $\overline{MU}_p=\overline{AF}_c$; $\overline{AF}_p=\overline{NM}_c$; $\overline{FA}_p=\overline{MN}_c$.

(3) θ_p^1 and θ_p^2 are in the same column; the letter immediately below θ_p^1 forms θ_c^1, the letter immediately below θ_p^2 forms θ_c^2. When either θ_p^1 or θ_p^2 is at the bottom of the column, the top letter in that column becomes its cipher equivalent. Examples: $\overline{MC}_p=\overline{CB}_c$; $\overline{AW}_p=\overline{TA}_c$; $\overline{WA}_p=\overline{AT}_c$; $\overline{QU}_p=\overline{YI}_c$.

(4) θ_p^1 and θ_p^2 are identical; they are to be separated by inserting a null, usually the letter X or Q, and subsequently enciphered by the pertinent rule from above. For example, the word BATTLES would be enciphered thus:

```
BA TX TL ES
DM RW CO KP
```

The Playfair square is automatically reciprocal so far as encipherments of type (1) above are concerned; but this is not true of encipherments of type (2) and (3).

e. It is not essential that the small matrices used for digraphic substitution be in the shape of perfect squares; rectangular designs will serve equally well, with little or no modification in procedure.[12] For example, each section of, say, a four-square matrix could be constructed with four rows containing six letters each by having U_p serve for V_p, as well as I_p for J_p. Furthermore, it is possible to expand the sections of a digraphic matrix to 28, 30, or more characters by the following subterfuge, without introducing digits or symbols into the cipher text.[13] One

[12] However, because the terms "four-square matrix", "two-square matrix", and "Playfair square" have become firmly fixed in cryptologic literature and practice, they continue to be applied to all such matrices, even when the "squares" of such matrices do not contain an equal number of rows and columns (that is, even when they are not square).

[13] The addition of any symbols such as the digits 1, 2, 3,... into a matrix solely to augment the number of elements to 27, 28, 30, 32, or 36 characters would not be considered practicable, since such a procedure would result in producing cryptograms containing intermixtures of letters and figures.

of the letters of the alphabet may be omitted from the set of 26 letters, and this letter may then be replaced by 2, 3, or more pairs of letters, each pair having as one of its members the omitted single letter. The 5x6 Playfair square of Figure 57a has been derived thus; the letter K has been omitted as a single letter, and the number of characters in the rectangle has been made a total of 30 by the addition of five combinations of K with other letters. An interesting consequence of this modification is that certain irregularities are introduced in any cryptogram produced through its use; for example, (1) occasionally a plaintext digraph is replaced by ciphertext trigraph or tetragraph, such as $\overline{AM}_p$=$\overline{HKU}_c$ and $\overline{EP}_p$=$\overline{KEKO}_c$; and (2) variant values may appear--$\overline{BKE}_c$, $\overline{DKE}_c$, $\overline{KEP}_c$, $\overline{GP}_c$, and $\overline{TP}_c$ all may be used to represent $\overline{CK}_p$. As far as the deciphering is concerned, there is no difficulty because any K occurring in the cipher text is considered as invariably forming a ligature with the succeeding letter, taking the pair of letters as a unit; and, when a plaintext unit is obtained containing one of the K-pairs, the letter after the K is disregarded; for example, $\overline{CKO}_p$ is read as CK. The four-square matrix in Fig. 57b has also been constructed using the foregoing

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

Figure 57a.

	B	2	E	5	R	L	A	B	C	D	E	F	
	I	9	N	A	1	C	G	H	I	J	KA	KE	
θ_p^1	3	D	4	F	6	G	KI	KO	KU	KY	L	M	θ_c^1
	7	H	8	J	Ø	K	N	O	P	QA	QE	QI	
	M	O	P	Q	S	T	QO	QU	QY	R	S	T	
	U	V	W	X	Y	Z	U	V	W	X	Y	Z	
	A	B	C	D	E	F	M	U	N	I	9	C	
	G	H	I	J	KA	KE	3	H	8	A	1	B	
θ_c^2	KI	KO	KU	KY	L	M	2	D	4	E	5	F	θ_p^2
	N	O	P	QA	QE	QI	6	G	7	J	Ø	K	
	QO	QU	QY	R	S	T	L	O	P	Q	R	S	
	U	V	W	X	Y	Z	T	V	W	X	Y	Z	

Figure 57b.

subterfuge. With this latter matrix, numbers in the plain text may be enciphered, still without producing cipher text containing numbers; for example, the plain text "HILL 3406" would be represented by the cipher QAB AT KUKI NQE, which would be regrouped into groups of five letters and sent as QABAT KUKIN QE...

f. Figure 58 shows a numerical four-square matrix which presents a rather interesting feature in that it makes possible the substitution of 3-figure combinations for digraphs in a unique manner. To encipher a message one proceeds as usual to find the numerical equivalents of a pair, and then these numbers are added together. Thus:

Plain text:	PR	OC	EE	DI	NG
	275	350	100	075	325
	9	13	24	18	7
Cipher text:	284	363	124	093	332

	A	B	C	D	E	000	025	050	075	100	
	F	G	H	I	K	125	150	175	200	225	
Sec. 1 (θ^1_p)	L	M	N	O	P	250	275	300	325	350	Sec. 3 (θ^1_c)
	Q	R	S	T	U	375	400	425	450	475	
	V	W	X	Y	Z	500	525	550	575	600	
	0	1	2	3	4	V	Q	L	F	A	
	5	6	7	8	9	W	R	M	G	B	
Sec. 4 (θ^2_c)	10	11	12	13	14	X	S	N	H	C	Sec. 2 (θ^2_p)
	15	16	17	18	19	Y	T	O	I	D	
	20	21	22	23	24	Z	U	P	K	E	

Figure 58.

In deciphering, the greatest multiple of 25 contained in the group of three digits is determined; then this multiple and its remainder are used to form the elements for determining the plaintext pair in the usual manner. Thus, 284=275+9=PR.

g. Thus far all the small-matrix methods have involved only digraphic substitution. The two matrices together illustrated in Figures 59a and b may be used to provide a system for encipherment which is partly trigraphic; the adverb "partly" has been used because this particular system will yield trigraphic encipherment approximately 88.5% of the time in ordinary text and digraphic encipherment approximately 11.5% of the time.[14] In this case the cipher equivalents of the trigraphs

[14] These figures are based on the number of trigraphs ending in one of the 15 highest-frequency letters (ETNROAISDLHCFPU), and on the number of trigraphs ending with other letters.

H_1	H_2	H_3	H_4	Y_1	Y_2	Y_3	Y_4	D_1	D_2	00	01	02	03	04	05	06	07	08	09
D_3	D_4	R_1	R_2	R_3	R_4	A_1	A_2	A_3	A_4	10	11	12	13	14	15	16	17	18	19
U_1	U_2	U_3	U_4	L_1	L_2	L_3	L_4	I_1	I_2	20	21	22	23	24	25	26	27	28	29
I_3	I_4	C_1	C_2	C_3	C_4	B_1	B_2	B_3	B_4	30	31	32	33	34	35	36	37	38	39
E_1	E_2	E_3	E_4	F_1	F_2	F_3	F_4	G_1	G_2	40	41	42	43	44	45	46	47	48	49
G_3	G_4	K_1	K_2	K_3	K_4	M_1	M_2	M_3	M_4	50	51	52	53	54	55	56	57	58	59
N_1	N_2	N_3	N_4	O_1	O_2	O_3	O_4	P_1	P_2	60	61	62	63	64	65	66	67	68	69
P_3	P_4	Q_1	Q_2	Q_3	Q_4	S_1	S_2	S_3	S_4	70	71	72	73	74	75	76	77	78	79
T_1	T_2	T_3	T_4	V_1	V_2	V_3	V_4	W_1	W_2	80	81	82	83	84	85	86	87	88	89
W_3	W_4	X_1	X_2	X_3	X_4	Z_1	Z_2	Z_3	Z_4	90	91	92	93	94	95	96	97	98	99
00	01	02	03	04	05	06	07	08	09	Q_1	Q_2	Q_3	Q_4	U_1	U_2	U_3	U_4	E_1	E_2
10	11	12	13	14	15	16	17	18	19	E_3	E_4	S_1	S_2	S_3	S_4	T_1	T_2	T_3	T_4
20	21	22	23	24	25	26	27	28	29	I_1	I_2	I_3	I_4	O_1	O_2	O_3	O_4	N_1	N_2
30	31	32	33	34	35	36	37	38	39	N_3	N_4	A_1	A_2	A_3	A_4	B_1	B_2	B_3	B_4
40	41	42	43	44	45	46	47	48	49	L_1	L_2	L_3	L_4	Y_1	Y_2	Y_3	Y_4	C_1	C_2
50	51	52	53	54	55	56	57	58	59	C_3	C_4	D_1	D_2	D_3	D_4	F_1	F_2	F_3	F_4
60	61	62	63	64	65	66	67	68	69	G_1	G_2	G_3	G_4	H_1	H_2	H_3	H_4	K_1	K_2
70	71	72	73	74	75	76	77	78	79	K_3	K_4	M_1	M_2	M_3	M_4	P_1	P_2	P_3	P_4
80	81	82	83	84	85	86	87	88	89	R_1	R_2	R_3	R_4	V_1	V_2	V_3	V_4	W_1	W_2
90	91	92	93	94	95	96	97	98	99	W_3	W_4	X_1	X_2	X_3	X_4	Z_1	Z_2	Z_3	Z_4

Figure 59a.

	1	2	3	4
1	-	E	T	N
2	R	O	A	I
3	S	D	L	H
4	C	F	P	U

Fig. 59b.

(or digraphs, as the case may be) are tetranomes. Encipherment is best illustrated by an example; this is given in the next subparagraph.

h. Let the text to be enciphered be a message beginning with the words "REFERRING TO YOUR MESSAGE NUMBER FIVE STOP ..." This is rewritten into trigraphs, with the proviso that the third letter of the trigraph be one of the letters contained in the small square in Fig. 59b; if the third letter is not one of these 15 letters, the plaintext grouping is left as a digraph; then the grouping into trigraphs (or digraphs) continues. Thus, the foregoing plain text would be written as follows:

REF ERR IN- GTO YOU RME SSA GEN UM- BER FI- VES TOP ...

In encipherment, it is to be noticed that R_p occurs four times in section 1 (as do all the letters) and E_p occurs four times in section 2; the proper combination of the 16 possibilities is determined by the coordinates of the third letter of the trigraph as indicated in the small square, Fig. 59b. Since the coordinates of F_p in this square are 42, then it is the 4th occurrence of R_p in section 1 and the 2d occurrence of E_p in section 2 which are used to obtain the equivalent for the trigraph $\overline{\text{REF}}_p$; this equivalent is 1905. When the plaintext unit as obtained above is only a digraph, it is the 1st occurrence of θ_p^1 which is used in section 1 and the 1st occurrence of θ_p^2 which is used in section 2; thus, "IN-" from the sample message beginning, above, would be enciphered 2828. The encipherment of the plaintext example above is then

REF	ERR	IN-	GTO	YOU	RME	SSA	GEN	UM-	BER	FI-	VES	TOP
1905	4081	2828	4719	0727	1372	7417	4118	2270	3807	4024	8806	8623

The cipher text could then be transmitted in groups of four digits, or, as a subterfuge to conceal the basic group length, the transmission could be in five-digit groups. In decipherment, the ciphertext tetranome is deciphered in the manner of the usual four-square matrix, and the location of the particular values for θ_p^1 and θ_p^2 will indicate the identity of the third plaintext letter, if any.

i. Now that the student has become familiar with the details of typical polygraphic substitution systems, he is ready to continue his cryptanalytic study with the treatment of methods for recognizing polygraphic substitution; these methods are described in the next paragraph.

67. Methods for recognizing polygraphic substitution.--a. The methods used to determine whether a given cryptogram is digraphic in character are usually rather simple. If there are many repetitions in a cryptogram or a set of cryptograms and yet the uniliteral frequency distribution gives no clear-cut indications of monoalphabeticity; if most of the repetitions contain an even number of letters and these repetitions for the most part begin on the odd letters and end on the even letters of the message, yet the cipher text does not yield to solution as a biliteral cipher when the procedures outlined in Sections VII and VIII are applied to it; if the cryptograms usually contain an even number of letters (exclusive of nulls); and if the cipher text is in letters and all 26 letters are not present and J or U are among the absent letters (or if the cipher is in digits and there is a limitation in the range of the text when divided into trinomes, this range usually being not greater than 001-676); then the encipherment may be assumed to be digraphic in nature.

b. Although the foregoing general remarks are true as far as they go, occasionally they may be difficult to apply with any clear-cut results unless a large volume of cipher text is available for study. To supplement them there are statistical tests which may be applied for the recognition of digraphic substitution. Just as the ϕ test and the Λ test may be applied to the uniliteral distribution of a cryptogram to help determine whether it is monoalphabetic with respect to single-letter plaintext units, so may these same tests be applied to the digraphic distribution of a cryptogram for the purpose of determining whether the cryptogram in question is monoalphabetic when considered as a digraphic cipher.

c. The basic form of the ϕ test is the same when applied to digraphic distributions as when applied to monographic--that is, uniliteral--distributions (see par. 27). It is only the plain and random constants that change, and "N" in the formulas now pertains to the number of digraphs under consideration, instead of the number of single letters.

To illustrate this, the formulas for computing the "digraphic phi plain (ϕ_p^2)" and the "digraphic phi random (ϕ_r^2)" are shown below:[15]

$$\phi_p^2 = .0069 \, N(N-1)$$

$$\phi_r^2 = .0015 \, N(N-1).$$

The "digraphic phi observed (ϕ_o^2)" is calculated in the usual manner, that is, by multiplying each f (which in this case is found in one of the cells of a digraphic distribution) by f-1, and then totalling all the values thus derived.

d. The Λ^2 test (or the "digraphic blank-expectation test") may be applied to a digraphic distribution just as easily as its monographic counterpart is applied to a uniliteral frequency distribution. For this purpose, Chart 8 is given below, showing the average number of blanks theoretically expected in digraphic distributions for plain text and for random text containing various numbers of digraphs (up to 200 digraphs). As can be seen, the chart contains two curves. The one labeled P applies to the average number of blanks theoretically expected in digraphic distributions based upon normal plaintext messages containing the indicated number of digraphs. The other curve, labeled R, applies to the average number of blanks theoretically expected in digraphic distri-

[15] The digraphic plain constant, .0069, was obtained by summing the squares of the probabilities of digraphs in English plain text; the digraphic random constant, .0015, is merely the decimal equivalent of 1/676. Further elaboration on the use of these constants, among others, will be given in Military Cryptanalysis, Part II.

~~RESTRICTED~~

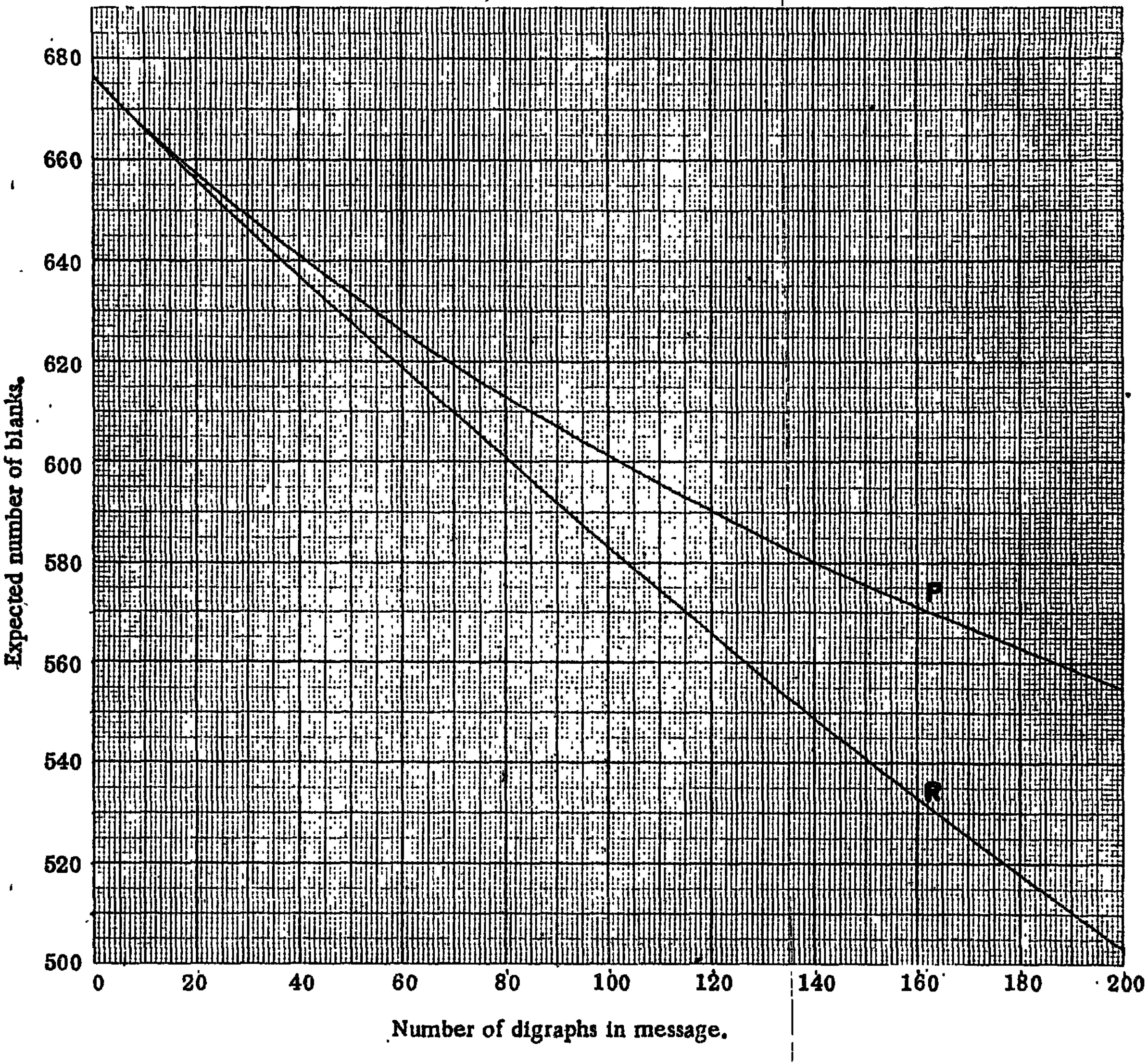


Chart 8.

butions based upon perfectly random assortments of digraphs. In using this chart one finds the point of intersection of the vertical coordinate corresponding to the number of digraphs in the message, with the horizontal coordinate corresponding to the observed number of blanks in the digraphic distribution for the message. If this point of intersection falls closer to curve P than it does to curve R, this is evidence

~~RESTRICTED~~

that the cryptogram is digraphic in nature[16]; if it falls closer to curve R than to curve P, this is evidence that the cryptogram is not digraphic in character.

e. Although it may not be necessary to resort to the use of the ϕ^2 and Λ^2 test to determine whether or not a particular cryptogram has been digraphically enciphered, it is well to know the application of these tests, since use has been made of them in difficult cases in operational practice. They may be helpfully employed in cases where the cryptanalyst is uncertain as to whether or not a single null has been added at the beginning of a cryptogram suspected to be a digraphic cipher; and these tests may also be found useful in the analysis of complex cases where the digraphic encipherment has been applied, not to adjacent letters of the plaintext message, but to digraphs composed of more-or-less separated letters in the message. Elaborations of these ideas will be treated in Military Cryptanalysis, Part II.

f. As for the recognition of trigraphic substitution ciphers--if most of the repetitions are a multiple of three letters in length, if these repetitions for the most part begin (when the cipher text is divided into trigraphs) with the first letters and end with the third letters of the trigraphs, and if the length of the cryptograms is for the most part a multiple of three letters, yet the cipher text does not yield to solution as a triliteral cipher, then the encipherment may be assumed to be trigraphic in nature.

g. Just as the ϕ test may be used as an aid in the recognition of digraphicity, it may theoretically be used for recognizing the trigraphic, tetragraphic, etc., nature of cryptograms, but its use for these latter purposes is much more limited because of the large amount of text which would be required to permit a valid application of the pertinent polygraphic ϕ test.

68. General procedure in the identification and analysis of polygraphic substitution ciphers.--a. Certain systems which at first glance seem to be polygraphic, in that groupings of plaintext letters are treated as units, are on closer inspection seen to be only partly polygraphic in character. Such is true of systems involving large tables of the type illustrated in Figs. 47a and b, and 48 (in par. 65, above),

[16] Unfortunately, such would also be the case if the cryptogram under consideration were a polyalphabetic cipher involving two alphabets. However, to distinguish between a digraphic cipher and a polyalphabetic cipher with two alphabets, a digraphic distribution could be made "off the cut", that is, made of those ciphertext digraphs which are formed by omitting the first letter of text and then dividing the remaining text into groups of two letters. If the system were digraphic, such a distribution would exhibit a poor ϕ_o^2; if the system were a two-alphabet substitution system, the ϕ_o^2 would be as satisfactory as that of the regular distribution, taken "on the cut".

wherein encipherment is by pairs but one of the letters in each pair is enciphered monoalphabetically, making these systems only pseudo-polygraphic. For example, using the table in Figure 48, any plaintext digraph beginning with "A" must be enciphered by a ciphertext digraph beginning with "W"; any plaintext digraph beginning with "B" must be enciphered by a ciphertext digraph beginning with "E"; etc. A cryptogram involving the use of this table may then be identified as such merely from a study of the uniliteral frequency distribution made on the initial letters of the cipher digraphs, since such a distribution would perforce be monoalphabetic.[17]

b. In certain other systems--namely, the four-square, two-square, and Playfair square systems of par. 66, above--the method of encipherment is by pairs, but the encipherments of the left-hand and right-hand members of the pairs show group relationships; this is not pseudo-polygraphic but, rather, partially-polygraphic. Cryptograms enciphered by means of systems of this latter type may not be readily identified as such merely through an examination of their cipher text, but their solution may be effected rather rapidly as soon as a few correct plaintext assumptions have been made therein. A more detailed treatment of this matter will be given in succeeding paragraphs of this section.

c. The analysis of cryptograms which have been produced by digraphic substitution is accomplished largely by the application of the simple principles of frequency of digraphs,[18] with the additional aid of digraphic idiomorphs and such special circumstances as may be known to or suspected by the cryptanalyst. The latter refer to peculiarities which may be the result of the particular method employed in obtaining the equivalents of the plaintext digraphs in the encrypting process, such as those mentioned in subpars. a and b, above. In general, if there is sufficient text to disclose the normal phenomena of repetition and idiomorphism, or if cribs are available to be used as an entering wedge, solution will be feasible. The foregoing general statements will be expanded upon in the following two subparagraphs, d and e.

d. When a digraphic system is employed in regular service, there is little doubt that traffic will rapidly accumulate to an amount more than sufficient to permit of solution by simple principles of frequency. Sometimes only two or three long messages, or a half-dozen of average length, are sufficient. For with the identification of only a few cipher

[17] For this purpose, the simplest and most economical way to obtain the uniliteral distributions for the initial and final letters of digraphs is to make a digraphic distribution and then add the tallies in each row to yield the distribution for the initial letters, and add the tallies in each column to obtain the distribution for the final letters.

[18] In this connection, it would be well for the student to familiarize himself with that portion of Appendix 2 which contains digraphic frequency data.

digraphs, larger portions of messages may be read because the skeletons of words formed from the few high-frequency digraphs very definitely limit the values that can be inserted for the intervening unidentified digraphs. For example, suppose that the plaintext digraphs RE, IN, ON, ND, NO, SI, NT, and TO are among those that have been identified by frequency considerations, corroborated by a tentatively identified long repetition; and suppose also that the enemy is known to be using a large table of 676 cells containing digraphs showing reciprocal equivalence between plaintext and ciphertext digraphs. Suppose the message begins as follows (in which the assumed values have been inserted):

XQ	VO	ZI	LK	AP	OL	ZX	PV	CK	IK	OL	UK	AT	HN	LK
		ND	IN		NT		RE			NT	NO			IN

VL	BN	OZ	BZ	DY	TY	LE	GI
	SI		ON	TO			

The initial words SECOND INFANTRY REGIMENT are readily recognized. Furthermore, if $\overline{CK}_c=\overline{GI}_p$, then $\overline{GI}_c=\overline{CK}_p$, which suggests ATTACK as the last word in the message beginning. This fragment of the message may now be completely recovered: SECOND INFANTRY REGIMENT NOT YET IN POSITION TO ATTACK.....

e. Just as the choice of probable words in the solution of uniliteral systems is aided or limited by the positions of repeated letters (see subpar. 49d), so, in digraphic ciphers, is the placing of cribs aided or limited by the positions of repeated digraphs. In this connection, several frequent words and phrases containing repeated digraphs have been tabulated for the student's aid, and this list of digraphic idiomorphs is presented as Section D in Appendix 3 (q.v.). Thus, if one is confronted by a ciphertext message containing the following repeated sequence (therefore likely to represent an entire word):

VI FW HM AZ FF FW RO

he may refer to the appropriate section of Appendix 3 which will disclose, on the basis of the idiomorphic pattern "AB -- -- -- AB" starting with the second cipher digraph, that the underlying plaintext word may be RE EN FO RC EM EN T, among others. Once a good start has been made and a few words have been solved, subsequent work is quite simple and straightforward. A knowledge of enemy correspondence, including data regarding its most common words and phrases, is of as much assistance in breaking down digraphic systems as it is in the solution of any other cryptosystems.

f. In the case of trigraphic substitution, analysis is made considerably more complex by the large amount of traffic required, not only for the initial entries, but also for further exploitation of the entering wedges. In effect, the solution of a trigraphic system closely parallels the solution of the syllabary portion of a large two-part code; these techniques will be discussed in Military Cryptanalysis, Part V.

69. Analysis of four-square matrix systems.--a. In all the small-matrix methods illustrated in paragraph 66, the encipherment is only partially digraphic because there are certain relationships between those plaintext digraphs which have common elements and their corresponding ciphertext digraphs, which will also have common elements. For example, in the four-square matrix given in Fig. 53, it will be noted that $\overline{AA}_p=\overline{FT}_c$, $\overline{AF}_p=\overline{FO}_c$, $\overline{AL}_p=\overline{FM}_c$, $\overline{AQ}_p=\overline{FL}_c$, and $\overline{AV}_p=\overline{FK}_c$. In each of these cases when A_p is the initial letter of the plaintext pair, the initial letter of the ciphertext equivalent is F_c. This, of course, is the direct result of the method; it means that the encipherment is monoalphabetic for the first half of each of these five plaintext pairs. This relationship holds true for four other groups of five pairs beginning with A_p; in effect, there are five cipher alphabets employed, not 25. Thus, this case differs from the case discussed under subpar. 68a only in that the monoalphabeticity is complete, not for half of all the pairs but only among the members of certain groups of pairs. In a true digraphic system, such as a system making use of a 676-cell randomized table, relationships of the foregoing type are entirely absent, and for this reason such a system is cryptographically more secure than small-matrix systems.

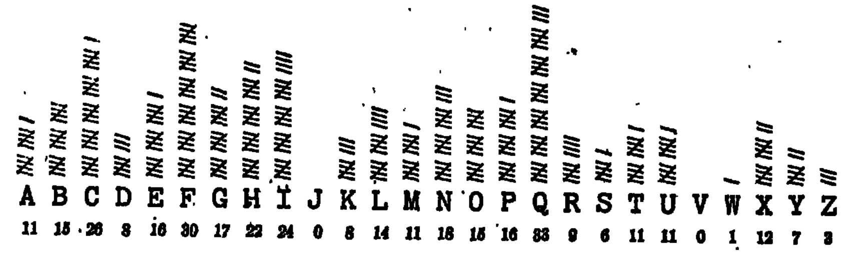
b. From the foregoing it is clear that when solution has progressed sufficiently to disclose a few values, the insertion of letters within the cells of the matrix to give the plaintext-ciphertext relationships indicated by the solved values immediately leads to the disclosure of additional values. Thus, the solution of only a few values soon leads to the breakdown of the entire matrix.

c. The following example will serve to illustrate the procedure. (1) Let the message be 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 27 28 29 30

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

(2) The cipher having been tested for standard alphabets (by the method of completing the plain-component sequence) and found to give negative results, a uniliteral frequency distribution is made. It is as follows:



(3) At first glance this may appear to the untrained eye to be a monoalphabetic frequency distribution, but upon closer inspection it is noted that, aside from the frequencies of four or five letters, the frequencies for the remaining letters are not very dissimilar. There are, in reality, no very marked crests and troughs--certainly not as many as would be expected in a monoalphabetic substitution cipher of equal length. The ϕ test, if taken (this test, as a rule, is not necessary with samples of text of sizes such as this), would show unsatisfactory results (ϕ_o=6084, as against ϕ_p=7870 and ϕ_r=4543).

(4) The message is carefully examined for repetitions of 4 or more letters, and all of them are listed:

	Frequency	Located in lines
TFHHAYZLQCIAAIQUCHTP (20 letters)........	2	H and L.
QMLFEQQIGOI (11 letters)................	2	C and F.
XIFBEX (6 letters)......................	2	C and D.
FEQQ....................................	3	C, D, F.
QMLF....................................	3	C, F, K.
BFHM....................................	2	B and G.
BSPK....................................	2	A and E.
GOIH....................................	2	D and K.

Since there are quite a few repetitions, two of considerable length, since all but one of them contain an even number of letters, since these repetitions with but two exceptions begin on odd letters and end on even letters, and since the message also contains an even number of letters (344), the cryptogram is retranscribed into 2-letter groups for further study. It is as follows:

Message transcribed in pairs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM	LF	EQ	QI
D.	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM	LF	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN
H.	CB	CQ	LQ	PN	FN	PN	IT	OR	TE	NC	CB	CN	TF	HH	AY
J.	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM	CB
K.	OY	CR	QQ	DP	RX	FN	QM	LF	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M.	ZL	QC	IA	AI	QU	CH	TP								

It is noted that all the repetitions listed above break up properly into digraphs except in one case, viz., FEQQ in lines C, D, and F. This latter seems rather strange, and at first thought one might suppose that a letter was dropped out or was added in the vicinity of the FEQQ in line D. But it may be assumed that the FE QQ in line D has no relation at all to the .F EQ Q. in lines C and F and is merely an accidental repetition.

(5) A digraphic distribution is made as follows:

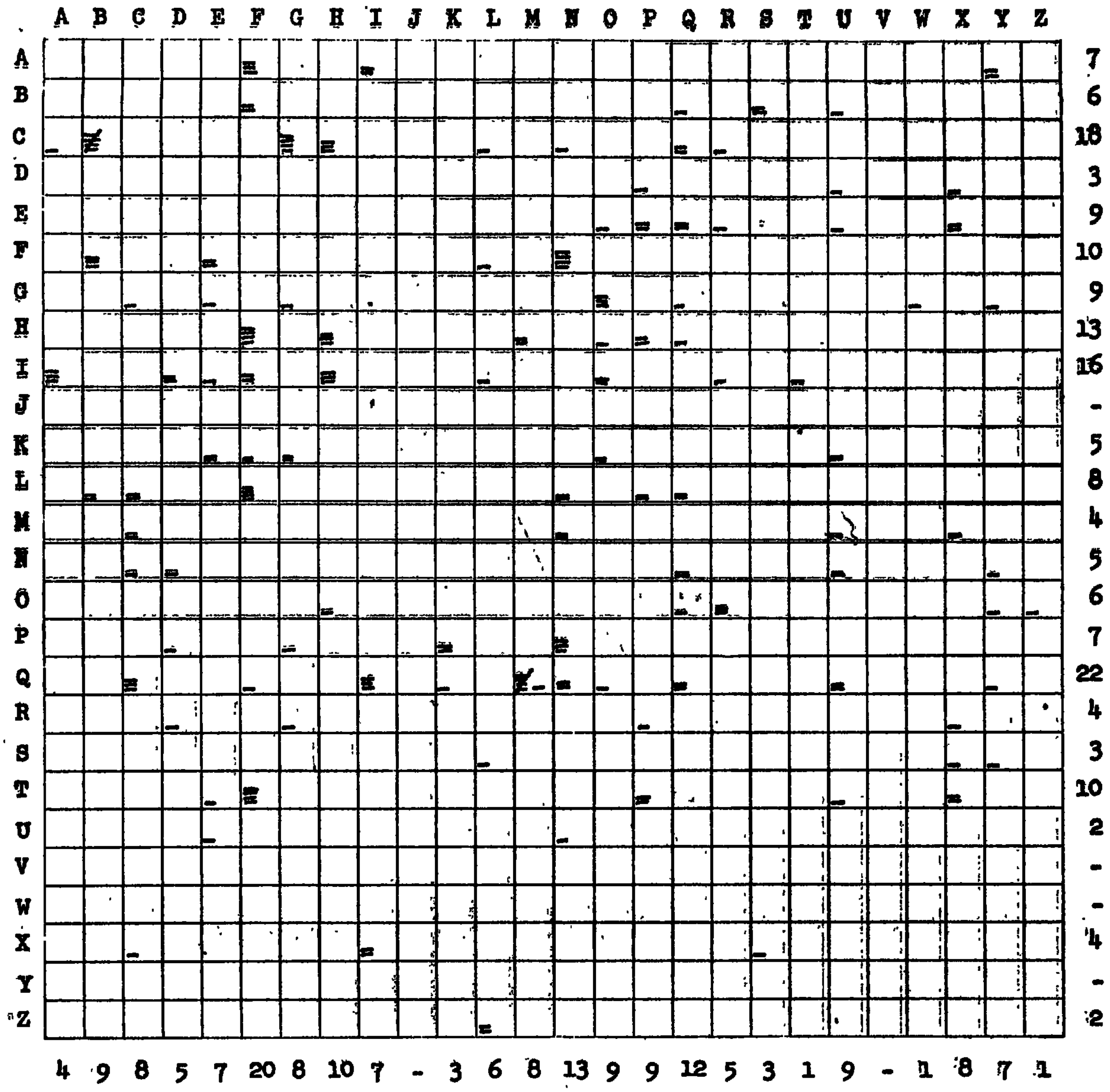


Figure 60.

(6) The appearance of the foregoing distribution for this message is quite characteristic of that for a digraphic substitution cipher. Although there are 676 possible digraphs, only 107 are present in the distribution; this parallels what is expected of normal plain text, since out of the 676 possible two-letter combinations (including "impossible plaintext digraphs" such as QQ, JK, etc., which might have been used for special indicators, punctuation marks, etc.) only about 300 are usually used in the construction of plain text.[19] The number of blank cells,

[19] The 300 most frequent digraphs comprise 95% of normal English plain text (Appendix 2, Table 7-A).

569, closely approximates the 565 which would be expected in a distribution made on a sample of plain text of this size, as shown by Chart 8. Furthermore, although there are many cases in which a digraph appears only once, there are quite a few in which a digraph appears two or three times, four cases in which a digraph appears four times, one case in which a digraph appears five times, and one in which a digraph appears six times. All of the foregoing observations concerning the distribution are reflected by the ϕ test: the observed digraphic phi value, 210, compares very favorably with the expected plain value (=.0069 x 172 x 171 = 203) as against the expected random value (=.0015 x 172 x 171 = 44). Thus all indications point to a *digraphic* substitution system.

(7) Since neither the ϕ_o (1780) and Λ_o (4) for the initial letters of the cipher digraphs nor the ϕ_o (1496) and Λ_o (2) for the final letters are too satisfactory in their approximation to the values expected for monoalphabetic distributions (ϕ_p=1962 and ϕ_r=1133; Λ_p=5 and Λ_r=0), the possibility of a *pseudo*-digraphic system is ruled out. There remain the possibilities of a *partially*-digraphic system employing a small matrix, or a *true* digraphic system employing a large, randomized table. In one common type of small-matrix system, the Playfair cipher, one of the telltale indications besides the absence of (usually) the letter J is the absence of cipher doublets, that is, two successive identical cipher letters. The occurrence of the double letters GG, HH, and QQ in the message under investigation eliminates the possibility of its being a normal Playfair cipher. For want of more accurate diagnostic criteria[20] *at this stage*,[21] the simplest thing to assume, from among the various hypotheses that remain to be considered, is that a four-square matrix is involved. One with normal alphabets (as being the simplest case) in Sections 1 and 2 is therefore set down (Figure 61*a*).

[20] Even a medical practitioner often cannot successfully diagnose a condition on the first visit. Cryptanalytically speaking, we are still on our "first visit". Subsequent probing will, we hope, reject or substantiate this or that hypothesis or assumption, until the patient (the cipher text) is recovered (i.e., brought back to plain text).

[21] However, see the treatment on the diagnosis of various types of digraphic systems in subpar 73*j*.

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

Figure 61a.

(8) The recurrence of the group QMLF, three times, and at intervals suggesting that it might be a sentence separator, leads to the assumption that it represents the word STOP. The letters Q, M, L, and F are therefore inserted in the appropriate cells in Sections 3 and 4 of the diagram. Thus (Fig. 61b):

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

Figure 61b.

These placements seem rather good from the standpoint that keyword-mixed sequences may have been used in these two sections. Moreover, in Section 3 the number of cells between L and Q is just one less than enough to contain all the letters M to P, inclusive; this suggests that one of these letters, probably N or O, is in the keyword portion of the sequence;

that is, near the top of Section 3. Without making a commitment in the matter, let us suppose that M follows L and that P precedes Q; then let both N and O, for the present, be inserted in the cell between M and P. Thus (Fig. 61c):

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

Figure 61c.

(9) Now, if the placement of P in Section 3 is correct, the cipher equivalent of $\overline{TH}_p$ will be $\overline{P\theta}_c$, and there should be a group of adequate frequency to correspond. Noting that $\overline{PN}_c$ occurs three times, it is assumed to represent $\overline{TH}_p$ and the letter N is inserted in the appropriate cell in Section 4. Thus (Fig. 61d):

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

Figure 61d.

(10) It is about time to try out these assumed values in the message. The proper insertions are made, with the following results:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM ST	LF OP	EQ	QI
D.	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM ST	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM ST	LF OP	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN TH
H.	CB	CQ	LQ	PN TH	FN	PN TH	IT	OR	TE	NC	CB	CN	TF	HH	AY
J.	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM ST	CB
K.	OY	CR	QQ	DP	RX	FN	QM ST	LF OP	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M.	ZL	QC	IA	AI	QU	CH	TP								

(11) So far no impossible combinations are in evidence. Beginning with group H4 in the message is seen the following sequence:

P N F N P N
T H . . T H

Assume it to be THAT THE. Then $\overline{AT}_p = \overline{\overline{FN}}_c$, and the letter N is to be inserted in row 4 column 1 of Section 4. But this is inconsistent with previous assumptions, since N in Section 4 has already been tentatively placed in row 2 column 4. Other assumptions for $\overline{\overline{FN}}_c$ are made: that it is, $\overline{IS}_p$ (THIS TH...); that it is $\overline{EN}_p$ (THEN TH...); but the same inconsistency is apparent. In fact the student will see that $\overline{\overline{FN}}_c$ must represent a digraph ending in F, G, H, I-J, or K, since N_c is tentatively located on the same line as these letters in Section 2. Now $\overline{\overline{FN}}_c$ occurs 4 times in the message. The digraph it represents must be one of the following:

DF, DG, DH, DI, DJ, DK
IF, IG, IH, II, IJ, IK
JF, JG, JH, JI, JJ, JK

OF, OG, OH, OI, OJ,
TK,
YF, YG, YH, YI, YJ, YK

Of these the only one likely to be repeated 4 times is OF, yielding

```
P N F N P N
T H O F T H which may be a part of
```

```
          C Q L Q P N F N P N I T       C Q L Q P N F N P N I T
          . N O R T H O F T H E .   or  . S O U T H O F T H E .
```

In either case, the position of the F in Section 3 is excellent: F . . . L in row 3. There are 3 cells intervening between F and L, into which G, H, I-J, and K may be inserted. It is not nearly so likely that G, H, and K are in the keyword as that I should be in it. Let it be assumed that this is the case, and let the letters G, H, and K be placed in the appropriate cells in Section 3. Thus (Fig. 61e):

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

Figure 61e.

Let the resultant derived values be checked against the frequency distribution. If the position of H in Section 3 is correct, then the digraph $\overline{ON}_p$, normally of high frequency, should be represented several times by $\overline{HF}_c$. Reference to Fig. 60 shows $\overline{HF}_c$ to have a frequency of 4. And $\overline{HM}_c$, with 2 occurrences, represents $\overline{NS}_p$. There is no need to go through all the possible corroborations.

```
                                         P N F N P N
   (12) Going back to the assumption that T H . . T H is part of the
expression
```

```
          C Q L Q P N F N P N I T       C Q L Q P N F N P N I T
          . N O R T H O F T H E .   or  . S O U T H O F T H E .,
```

it is seen at once from Fig. 61e that the latter is apparently correct and not the former, because $\overline{LQ}_c$ equals $\overline{OU}_p$ and not $\overline{OR}_p$. If $\overline{OS}_p$=$\overline{CQ}_c$, this means that the letter C of the digraph $\overline{CQ}_c$ must be placed in row 1 column 3 or row 2 column 3 of Section 3. Now the digraph $\overline{CB}_c$ occurs 5 times, $\overline{CG}_c$, 4 times, $\overline{CH}_c$, 3 times, $\overline{CQ}_c$, 2 times. Let an attempt be made to deduce the exact position of C in Section 3 and the positions of B, G, and H in Section 4. Since F is already placed in Section 4, assume G and H directly follow it, and that B comes before it. How much before? Suppose a trial be made. Thus (Fig. 61f):

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

Figure 61f.

By referring now to the frequency distribution, Fig. 60, after a very few minutes of experimentation it becomes apparent that the following is correct:

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

Figure 61g.

(13) The identifications given by these placements are inserted in the text, and solution is very rapidly completed. The final matrix and deciphered text are given below.

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

Figure 61h.

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

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

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

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

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

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

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

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

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

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

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

M. Z L Q C I   A A I Q U   C H T P
   Z E R O E   I G H T D   A S H A
```

d. In the solution of four-square cryptograms, advantage may be taken not only of the general type of digraphic idiomorphs mentioned in subpar. 68e, above, but also of a special type of partial idiomorphism present in any four-square cryptograms involving the use of a matrix in which the plain components consist of normal alphabets normally inscribed.[22] As an illustration, let the digraphs $\overline{SO}$ $\overline{UT}$ (H.) be enciphered by means of any four-square having normal alphabets in Sections 1 and 2, and it will be found that in the encipherment the initial letter of the cipher digraph representing $\overline{SO}_p$ will be identical to the initial letter of the cipher digraph representing $\overline{UT}_p$, regardless of how the cipher components are constructed. On this basis, a brief list of specialized single-letter patterns have been compiled for use in the solution of such a digraphic system; this list of "four-square digraphic idiomorphs" constitutes Section F of Appendix 3.

e. It is interesting to note how much simpler the technique of analysis is in the case of so-called inverse four-square ciphers, which involve the use of a matrix wherein the ciphertext sections contain normal alphabets, the plain components being mixed. For example, referring to Fig. 53, suppose that Sections 3 and 4 are used as the source of the plaintext pairs, and Sections 1 and 2 as the source of the ciphertext pairs; then $ON_p=ET_c$, $EH_p=GE_c$, etc. The simplicity of the analytic procedure will be made clear by the following exposition.

(1) To solve a message enciphered with an inverse four-square matrix, it is necessary to perform two steps. First, convert the ciphertext pairs into their plain-component equivalents by "deciphering" the message with a matrix in which all four sections contain normal alphabets; this operation yields two uniliteral substitution "ciphers", one composed of the odd letters, the other of the even letters. The second step is to solve these two monoalphabetic portions.

(2) As an example, let us consider the following cipher text, known (or assumed) to have been encrypted with a trinome-digraphic[23] system

[22] If any other known plain components were involved, the procedure of deriving a list of idiomorphic patterns would be modified to fit the particular case.

[23] If the cipher text were being examined "from cryptanalytic scratch", the limitations (003-595) of the cipher text when the latter is divided into trinomes for examination would have at once indicated that this grouping is the one which merits detailed analysis. The ϕ^2 test would then give an indication of the digraphic nature of the underlying cryptographic treatment.

incorporating a four-square matrix similar to that illustrated in Fig. 58, except that the plain-component sections have been changed:

```
20323   85081   83450   27934   11503   09168
27835   41804   50413   27416   33091   01092
20805   74135   35473   32626   91160   03218
46818   33930   91393   41104   41331   17296
24302   83832   28359   38022   61043   69130
15313   61041   00144   10101   82403   36168
46536   62663   44007   18345   01402   88152
47821   73933   81193   47924   04032   41306
08703   70914   19391   11607   71371   53595
00741   33381   33593   39340   63531   88133
```

(3) The first thing to be done is to construct a four-square matrix with the known ciphertext sections, and inscribe arbitrary alphabets in the plaintext sections, as follows:

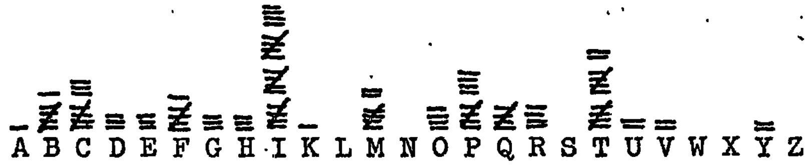
A	B	C	D	E	000	025	050	075	100
F	G	H	I	K	125	150	175	200	225
L	M	N	O	P	250	275	300	325	350
Q	R	S	T	U	375	400	425	450	475
V	W	X	Y	Z	500	525	550	575	600
Ø	1	2	3	4	A	B	C	D	E
5	6	7	8	9	F	G	H	I	K
10	11	12	13	14	L	M	N	O	P
15	16	17	18	19	Q	R	S	T	U
20	21	22	23	24	V	W	X	Y	Z

(4) The cipher text is then written in trinomes, and these trinomes are "deciphered" by means of the foregoing matrix, yielding the converted cipher text as follows:

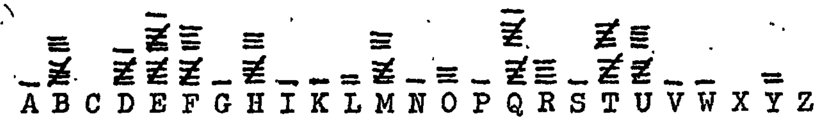
					5.					10					15
A	203 ID	238 IP	508 YF	183 IH	450 QD	279 PB	341 MT	150 FB	309 PH	168 IR	278 OB	354 PE	180 FH	450 QD	413 TM
B	274 PV	163 IM	309 PH	101 BE	092 CT	208 II	057 CH	413 TM	535 VM	473 TY	326 MD	269 PQ	116 BU	003 DA	218 IT
C	468 TT	183 IH	393 TQ	091 BT	393 TQ	411 RM	044 ER	133 IF	117 CU	296 MW	243 IU	028 DB	383 TF	228 IE	359 PK
D	380 QF	226 GE	104 EE	369 PU	130 FF	153 IB	136 GL	104 EE	100 AE	144 KQ	101 BE	018 DQ	240 FU	336 MO	168 IR
E	465 QT	366 MU	266 MQ	344 PT	007 CF	183 IH	450 QD	140 FQ	288 OM	152 HB	478 TE	217 HT	393 TQ	381 RF	193 IS
F	479 UE	240 FU	403 TB	241 GU	306 MH	087 CO	037 CM	091 BT	419 UR	391 RQ	116 BU	077 CD	137 HL	153 IB	595 VY
G	007 CF	413 TM	338 OO	133 IF	593 YT	393 TQ	406 RG	353 OE	188 IN	133 IF					

The distributions of the letters constituting the initial letters and final letters of the converted digraphs are as follows:

(Initial Letters)



(Final Letters)



(5) Using straightforward principles of frequency and partial idiomorphs,[24] the plain text (beginning with the opening words ENEMY RECONNAISSANCE...) is recovered, and the following equivalents are obtained for the converted cipher letters of the two alphabets:

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

[24] Note the ABA pattern of the first word in the message (ENEMY), made patent by the two-alphabet conversion process. Also note the 3-fold repetition (representing the plaintext word STOP) which, although hidden in the original cipher text, now comes to light.

Keyword-mixed sequences directly manifest themselves because the original enciphering matrix contained such sequences in Sections 1 and 2, inscribed in the same manner as were the arbitrary A-Z sequences which were used for the conversion. In fact, the key words of the two distributions might have been recovered from an analysis of the "profiles" of the distributions above, as described in subpar. 54<u>e</u>.

(6) The original enciphering matrix is then reconstructed, thus:

B	R	A	H	M	000	025	050	075	100
S	C	D	E	F	125	150	175	200	225
G	I	K	L	N	250	275	300	325	350
O	P	Q	T	U	375	400	425	450	475
V	W	X	Y	Z	500	525	550	575	600
Ø	1	2	3	4	W	A	G	N	E
5	6	7	8	9	R	B	C	D	F
10	11	12	13	14	H	I	K	L	M
15	16	17	18	19	O	P	Q	S	T
20	21	22	23	24	U	V	X	Y	Z

(7) Although the example illustrated was that of a numerical digraphic system, it is obvious that this technique of solution also applies to <u>literal</u> four-square systems in which the cipher components are known sequences. It should be clear to the student the tremendous difference it makes when it is possible to convert a <u>digraphic</u> system into a <u>two-alphabet</u> system; in a digraphic system, we are plagued by a potential 676 different elements in the cipher, whereas in a two-alphabet system we still have only 26 elements (in each of two sets, it is true) in the cipher text to be solved. This principle of conversion of cipher text into a secondary cipher text has application in some of the most complex types of cryptosystems; the student would do well to keep this in mind.

(8) As a further observation on inverse four-square systems, it is pointed out that where the <u>same</u> mixed alphabet is present in Sections 3 and 4, the problem is still easier, since the letters resulting from the conversion into plain-component equivalents all belong to the same, <u>single</u> mixed alphabet; thus such a digraphic system is reduced to an ordinary simple substitution cipher.

<u>f</u>. The solution of cryptograms enciphered by other types of small matrices is accomplished along lines very similar to those set forth in subparagraph <u>c</u> on the solution of a four-square cipher; this will be illustrated in subsequent paragraphs. There are, unfortunately, few means or tests which can be applied to determine in the early stages of the analysis <u>exactly</u> what type of digraphic system is involved in the first case under study. The author freely admits that the solution outlined in subparagraph <u>c</u> is quite artificial in that nothing is demonstrated in step (7) that obviously leads to or warrants the assumption that a four-square matrix is involved. The point was passed over with the quite bald statement that this was "from among the various hypotheses that remain to be considered"--and then the solution proceeded exactly as though this mere hypothesis had been definitely established. For example, the very first

results obtained were based upon assuming that a certain 4-letter repetition represented the word STOP and immediately inserting certain letters in appropriate cells in a four-square matrix with normal sequences in Sections 1 and 2. Several more assumptions were built on top of that, and very rapid strides were made. What if it had not been a four-square matrix at all? What if it had been some other type of not readily identifiable digraphic system? The only defense that can be made of what may seem to the student to be purely arbitrary procedure based upon the author's advance information or knowledge is the following: In the first place, in order to avoid making the explanation a too-long-drawn-out affair, it is necessary (and pedagogical experience warrants) that certain alternative hypotheses be passed over in silence. In the second place, it may now be added, after the principles and procedure have been elucidated (which at this stage is the primary object of this text) that if good results do not follow from a first hypothesis, the only thing the cryptanalyst can do is to reject that hypothesis and formulate a second hypothesis. In actual practice he may have to reject a second, third, fourth, ...nth hypothesis. In the end he may strike the right one--or he may not. There is no assurance of success in the matter. In the third place, one of the objects of this text is to show how certain cryptosystems, if employed for military purposes, can readily be broken down. Assuming that some type of digraphic system is in use, and that daily changes in key words are made, it is possible that the traffic of the first day might give considerable difficulty in solution if the specific type of digraphic system were not known to the cryptanalyst. But by the time two or three days' traffic had accumulated it would be easy to solve, because probably by that time the cryptanalytic personnel would have successfully analyzed the cryptosystem and thus learned what type of matrix or table the enemy is using.

70. Analysis of two-square matrix systems.--a. Cryptosystems involving either vertical two-square or horizontal two-square matrices may be identified as such and solved by capitalizing on the cryptographic peculiarities and idiosyncracies of these systems. It will be noted that, considering the mechanics of the cryptosystems, in vertical two-square matrices employing the normal enciphering conventions,[25] exactly 20% of the 625 "possible" plaintext digraphs will be "transparent" (i.e., self-enciphered) in cipher text; in horizontal two-square systems, exactly 20% of the 625 digraphs will be characterized by an "inverse transparency"

[25] That is, for vertical two-square systems, digraphs are self-enciphered if θ_p^1 and θ_p^2 fall in the same column in the matrix; and, for horizontal two-square systems, if θ_p^1 and θ_p^2 are in the same row, the ciphertext digraphs are the reversed plaintext digraphs.

(i.e., enciphered by the same digraphs reversed).[26] Therefore, if an examination of a cryptogram or a set of cryptograms discloses a goodly portion of what appear to be direct transparencies (cipher digraphs which could well be plaintext digraphs), it may then be assumed that a vertical two-square matrix has been used for the encryption. On the other hand, if a large number of cipher digraphs could be "good" plaintext digraphs if the positions of the letters were reversed, then it may be assumed that the cryptosystem involved a horizontal two-square matrix. Sometimes skeletons of words or even of whole phrases are self-evident in such cipher text, thus affording an easy entering wedge into the cryptosystem.

b. An example will best serve to illustrate the techniques of identification and subsequent solution of a two-square matrix cipher. The following naval message is to be studied:

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

Preliminary steps in analysis are made according to the procedures already described in this text, and the hypothesis of monographic, uniliteral encipherment (with either standard or mixed cipher alphabets) has been rejected. Multiliteral substitution, or digraphic substitution, comes next

[26] Although 625 "possible" plaintext digraphs are involved, the identity of digraphs actually used in plain text limit this figure considerably. Furthermore, the frequencies of the plaintext digraphs actually used come into consideration, in conjunction with the location of the letters of these digraphs in any particular two-square matrix. Thus, from the cryptanalyst's standpoint, there are "excellent" two-square matrices giving a high self-encipherment rate for high frequency plaintext digraphs, and there are "poor" two-square matrices which have a potentially high self-encipherment rate only for those low frequency plaintext digraphs which may not occur at all in a given cryptogram.

into consideration. The cipher text is written in digraphs, as follows:

				5					10					15
A	UO	DL	CE	NO	AN	SI	GL	BB	EI	RI	RC	RG	LN	MO	LC
B	PT	ER	GR	BB	OE	GP	AB	QW	NN	KS	IP	CR	MO	OR	AP
C	DE	AM	HA	NX	RA	IE	DA	IR	MA	GB	EK	HS	LC	DD	LC
D	TQ	OR	EN	DT	MD	TI	AQ	FI	EQ	TA	NN	BF	NO	UO	OS
E	SN	NN	RK	TA	SE	SN	HL	PO	NN	KS	IP	CR	CE	NO	IS
F	HL	IR	KP	LO	NO	NZ	UC	TA	LT	OI	IH	OC	NO	CE	RA
G	OS	DI	NO	EE	KR	LC	UB	RA	OS	DI	IP	DA	RC	OG	GR
H	OL	NO	CW	DI	LP	OI	LN	QX	DI	GL	RB	BQ	YF	SS	RA
J	VY	OI	GR	SL	XX										

Figure 62.

Noting the 8-letter repetition 90 letters apart, the 6-letter repetition 16 letters apart, and the 4-letter repetition at an interval of 220 letters, and that those repetitions begin on odd letters and end on even letters, credence is given to the grouping of the cipher text into pairs of letters. A digraphic distribution is then made, illustrated in Fig. 63.

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

Figure 63.

c. The ϕ_o^2, 152, is most satisfactory when compared with ϕ_p^2 (107) and ϕ_r^2 (23). Since the cryptogram has all the earmarks of a digraphic cipher, and no manifestations are found to support the hypothesis of a multiliteral system, the next problem is the specific determination of the particular kind of digraphic system involved. It may be noted that there are quite a few digraphs in the cipher text which resemble good plaintext digraphs, proportionally more so than, for instance, in the cryptogram in subpar. 69c; the cryptologic finger points to the possibility of a two-square system. However, since the words "good digraphs" are semantically elusive, let us attempt to determine statistically whether or not a two-square system might be involved and, if a two-square, whether it is more probably a vertical or a horizontal two-square.[27]

d. First, for the purpose of determining whether "direct transparencies" or "inverse transparencies" predominate in this cryptogram, the digraphs of the distribution in Fig. 63 will be set down in tabular form, with an indication of their frequency in the cryptogram, and with data relative to the probability of these digraphs as plaintext digraphs, and as plaintext digraphs when reversed. In the table on page 194, col. (1) is a listing of the ciphertext digraphs; col. (2) is the frequency of the ciphertext digraph as it occurs in the cryptogram; col. (3) is the logarithm of the theoretical plaintext frequency of the particular digraph (from Table 15, Appendix 2); col. (4) represents the products of the entries in cols. (2) and (3); col. (5) is the logarithm of the theoretical plaintext frequency of the reversed digraph (from Table 15, Appendix 2); and col. (6) represents the products of the entries in cols. (2) and (5). From this, the sum of the values in col. (4), 58.34, is taken to be the "direct transparency" value, and the sum of the values in col. (6), 63.02, is taken to be the "inverse transparency" value. Thus, since this particular cryptogram has an "inverse transparency" value which is higher

[27] The test to be described in the following subparagraphs is based on an evaluation of those instances wherein the observed frequency of any particular ciphertext digraph approximates the frequency with which the particular digraph, or its reversal, would be expected to occur if considered as a plaintext digraph. Any such correlation which occurs in a four-square or Playfair cipher, or in a cryptogram produced by a large randomized digraphic table, is purely accidental because it is not a result of the mechanics of the system. However, in two-square cryptograms such correlation is caused by the mechanics of the system in the encipherment of 20% of the possible plaintext digraphs, and these causal instances of correlation occur in addition to any accidental instances which may arise in the encipherment of the remaining 80%. Thus, if a digraphic cipher exhibits merely the random expectation of correlation both when the particular ciphertext digraphs are considered as they are and when their reversals are considered, the cryptogram may be assumed to involve a system other than two-square. If a digraphic cipher exhibits more than the random expectation of correlation, either when the particular digraphs are considered direct or when considered reversed, it may be assumed to involve two-square encipherment; and the particular consideration--that of the digraphs direct or that of the digraphs reversed--which gives rise to the greater degree of correlation indicates whether the cryptogram involves a vertical two-square or a horizontal two-square, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
AB	1	.45	0.45	.38	0.38
AM	1	.61	0.61	.78	0.78
AN	1	.89	0.89	.72	0.72
AP	1	.58	0.58	.61	0.61
AQ	1	.00	0.00	.00	0.00
BB	2	.00	0.00	.00	0.00
BF	1	.00	0.00	.00	0.00
BQ	1	[illegible]	0.00	.00	0.00
CE	3	.76	2.28	.76	2.28
CR	2	.38	0.76	.53	1.06
CW	1	.13	0.13	.00	0.00
DA	2	.76	1.52	.73	1.46
DD	1	.51	0.51	.51	0.51
DE	1	.77	0.77	.88	0.88
DI	4	.73	2.92	.45	1.80
DL	1	.33	0.33	.53	0.53
DT	1	.62	0.62	.45	0.45
EE	1	.81	0.81	.81	0.81
EI	1	.73	0.73	.59	0.59
EK	1	.00	0.00	.45	0.45
EN	1	.99	0.99	.87	0.87
EQ	1	.58	0.58	.00	0.00
ER	1	.94	0.94	.96	0.96
FI	1	.80	0.80	.55	0.55
GB	1	.00	0.00	.00	0.00
GL	2	.25	0.50	.13	0.26
GP	1	.25	0.25	.00	0.00
GR	3	.42	1.26	.48	1.44
HA	1	.67	0.67	.25	0.25
HL	2	.13	0.26	.13	0.26
HS	1	.38	0.38	.72	0.72
IE	1	.59	0.59	.73	0.73
IH	1	.00	0.00	.77	0.77
IP	3	.48	1.44	.45	1.35
IR	2	.73	1.46	.75	1.50
IS	1	.78	0.78	.77	0.77
KP	1	.00	0.00	.00	0.00
KR	1	.00	0.00	.13	0.13
KS	2	.13	0.26	.13	0.26
LC	4	.33	1.32	.42	1.68
LN	2	.13	0.26	.42	0.84
LO	1	.59	0.59	.67	0.67
LP	1	.33	0.33	.59	0.59
LT	1	.51	0.51	.42	0.42
MA	1	.78	0.78	.61	0.61
MD	1	.13	0.13	.42	0.42
MO	2	.55	1.10	.72	1.44
NN	4	.51	2.04	.51	2.04
NO	7	.66	4.62	.92	5.74
NX	1	.00	0.00	.13	0.13
NZ	1	.00	0.00	.00	0.00
OC	1	.51	0.51	.80	0.80
OE	1	.33	0.33	.58	0.58
OG	1	.25	0.25	.45	0.45
OI	3	.42	1.26	.80	2.40
OL	1	.67	0.67	.59	0.59
OR	2	.89	1.78	.74	1.48
OS	3	.61	1.83	.62	1.86
PO	1	.64	0.64	.72	0.72
PT	1	.51	0.51	.25	0.25
QW	1	.00	0.00	.00	0.00
QX	1	.00	0.00	.00	0.00
RA	4	.80	3.20	.82	3.28
RB	1	.25	0.25	.25	0.25
RC	2	.53	1.06	.38	0.76
RG	1	.48	0.48	.42	0.42
RI	1	.75	0.75	.73	0.73
RK	1	.13	0.13	.00	0.00
SE	1	.84	0.84	.86	0.86
SI	1	.77	0.77	.78	0.78
SL	1	.25	0.25	.45	0.45
SN	2	.38	0.76	.71	1.42
SS	1	.67	0.67	.67	0.67
TA	3	.74	2.22	.83	2.49
TI	1	.82	0.82	.73	0.73
TQ	1	.13	0.13	.00	0.00
UB	1	.33	0.33	.25	0.25
UC	1	.33	0.33	.38	0.38
UO	2	.13	0.26	.79	1.58
VY	1	.00	0.00	.00	0.00
XX	1	.00	0.00	.00	0.00
YF	1	.56	0.56	.13	0.13
	125		58.34		63.02

(1) Identity of cipher digraph appearing in the cryptogram.
(2) Frequency of the particular digraph as it occurs in the cryptogram.
(3) Logarithm of theoretical plaintext frequency of the particular digraph (from Table 15, Appendix 2).
(4) Product of entries in columns (2) and (3).
(5) Logarithm of theoretical plaintext frequency of the digraph's reversal (from Table 15, Appendix 2).
(6) Product of entries in columns (2) and (5).

than the "direct transparency" value, it may be assumed[28] to involve a horizontal two-square--if, indeed, two-square encipherment has been employed. It is now for us to establish whether or not this latter is the case, and this will be done by determining whether or not the foregoing observed value, 63.02, is representative of the degree of transparency which may be expected in a horizontal two-square cipher. (If the "direct transparency" value had been the higher of the two, then it would have been more probable that a vertical two-square were involved, and it would be necessary to determine whether or not this observed value was representative of the degree of transparency expected in a vertical two-square cipher).

e. The observed "inverse transparency" value (selected in this case because it is the higher observed value) will be compared with the value expected from a horizontal two-square cryptogram of the same size, and if this observed value is as great as or greater than the transparency value expected for horizontal two-squares, the cryptogram may be considered to be a horizontal two-square cipher; if the observed value is lower than the expected two-square value, decision will have to be suspended.[29] The transparency value expected in a horizontal two-square cipher containing N digraphs is computed by multiplying N by .3388, which in this case

[28] Actually, if the two-square hypothesis is made, the difference between the horizontal two-square value and the vertical two-square value will indicate the degree of probability of the higher score over the lower. In this case, the difference of 4.68 (= 63.02 - 58.34), which represents a difference of log scores, is equivalent to an overwhelming ratio of 100 billion to 1 (i.e., $224^{4.68}$ to 1) in favor of the hypothesis of a horizontal two-square. The foregoing computation involves an aspect of mathematics which will be given detailed treatment in Military Cryptanalysis, Part III.

[29] For the benefit of the student with a background in statistics, it is pointed out that by abiding by the stipulation "as great or greater", some cryptograms which actually are the result of two-square encipherment may be rejected by this stipulation, but it will insure that only a relatively few non-two-square cryptograms will be accepted. A better approach of a statistical nature would involve, first, computing the expected value for non-two-squares as well as that for two-squares. Then, any observed value falling below the expected two-square value could be expressed in terms of the number of standard deviations (i.e., the sigmage) from this expected two-square value and from the expected non-two-square value. Finally, the particular expected value which would be considered as significant would be the one from which the observed value differed by the smaller number of standard deviations. The concept of standard deviation will be treated in Military Cryptanalysis, Part III.

yields 42.35 (= .3388 x 125).[30] The observed value for the cryptogram, 63.02, is much higher than the expected value, 42.35. Thus, it has been proven statistically that the cryptogram at hand involves two-square encipherment, particularly, horizontal two-square encipherment.

f. Having now proved that the cryptogram at hand is a horizontal two-square cipher, the next step is to assume some plain text in the message, guided by probable inverse transparencies (inverse because the system has been identified as a horizontal two-square) in the cipher text. Referring to the work sheet in Fig. 62, the repeated sequence at B9 and E9 is assumed to represent the plain text TA SK FO RC (E-), on the basis of $\overline{KB}_c=\overline{SK}_p$, and $\overline{CR}_c=\overline{RC}_p$. The plaintext-ciphertext values are now

[30] In the case of vertical two-squares, N would be multiplied by the constant .3610. The mathematical considerations underlying this test and their proofs (involving Bayes' theorem and Bayes' factors) are beyond the scope this text; however, for the benefit of the mathematician, the derivation of the foregoing constants is explained below, along with the derivation of the constant used for computing the expected transparency value for non-two-squares. In the formulas, below,

$\sum_{AB}$ = the summation over all digraphs AA-ZZ

F_{AB} = the frequency of a given digraph AB as found in Table 6A, Appendix 2

α_{AB} = the logarithm (to the base 224) of the frequency of a given digraph AB as found in Table 15, Appendix 2

For vertical two-squares,

$$k = \sum_{AB} \alpha_{AB} \left[.80(.0015) + \frac{.20\, F_{AB}}{5000} \right] = .3610$$

For horizontal two-squares,

$$k = \sum_{AB} \alpha_{BA} \left[.80(.0015) + \frac{.20\, F_{AB}}{5000} \right] = .3388$$

For non-two square digraphic systems,

$$k = \frac{\alpha_{AB}}{676} = .2737$$

recorded[31] in a skeleton reconstructibn diagram as illustrated in Fig. 64a. At A3, the assumption of (-R) EC ON NA IS SA NC (E-) is tossed off without much ado, since four of the six digraphs concerned are transparent. The plain-cipher relationships from this assumption are added to the reconstruction diagram, as shown in Fig. 64b. Continuing in this vein, the plain text (-A) IR CR AF (T-) is inserted at A1Ø, and the plain text (-B) AT TL ES HI (P-) is inserted at J3; the successive cumulative reconstruction diagrams for these two assumptions are shown in Figs. 64c

			T	N			
			N	A			
		S			K		
	F					I	
	P					O	
R							C

Figure 64a.

	O					T	N				
					L	N	A			B	
					S		G	K			
				F					I		
				P					O		
		E	R			B				C	
I											S

Figure 64b.

		O					T	N					
A	C		I			L	N	A			B	S	R
						S		G	K				
					F					I			
					P					O			
			E	R			B	E			C		
G													F

Figure 64c.

[31] During the reconstruction of the squares of the matrix, the student should keep clear in his skeleton diagram which letters are in the same row, and which are in the same column. It will be found expeditious to draw a dividing line (either horizontal or vertical, depending on the type of two-square matrix involved) on the page to keep the elements of the two squares independent, recording the values which are in the same row or column and writing down the letters as they are assumed. In the early stages of this process the student must exercise care in recording the letters so that no false relationships are formed; in other words, the values should be written down so that they are not in the same row or column with any letters other than those with which they are known to be related. This will entail spreading the work rather widely over the page initially, then gradually telescoping and reducing the size of the reconstruction diagram as the work progresses, until in the end it will be reduced to a concise matrix of two 5x5 squares.

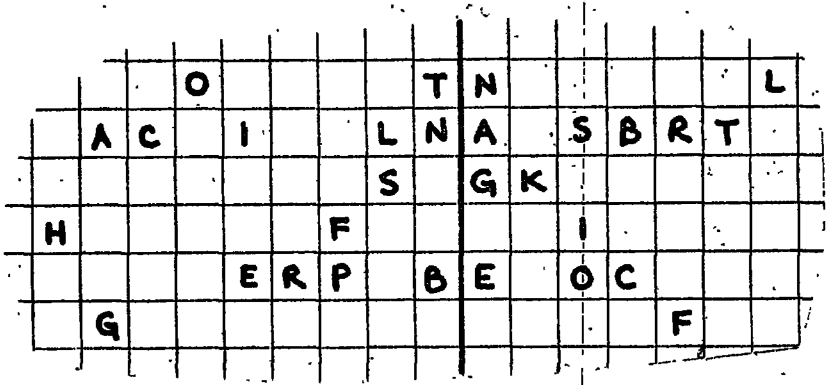


Figure 64d.

and 64d below. It is to be noted that at J7, $\overline{OC}_c=\overline{P\theta}_p$; but since in Fig. 64d it has already been determined that $OC_c=\theta S_p$, then $\overline{OC}_c$ must equal $\overline{PS}_I$ aking the word BATTLESHIPS rather than in the singular.

g. At this point the partially filled-in work sheet will look as follows:

					5					10					15
A	UO	DL -R	CE EC	NO ON	AN NA	SI IS	GL SA	BB NC	EI EA	RI IR	RC CR	RG AF	LN T-	MO E-	LC
B	PT	ER RE	GR -E	BB NC	OE EO	GP -E	AB NE	QW	NN TA	KS SK	IP FO	CR RC	MO E-	OR RO	AP -E
C	DE	AM	HA	NX	RA AR	IE -O	DA	IR -O	MA	GB -E	EK	HS	LC	DD	LC
D	TQ	OR RO	EN BA	DT	MD	TI IT	AQ	FI -R	EQ	TA AT	NN TA	BF	NO ON	UO	OS
E	SN NS	NN TA	RK	TA AT	SE IO	SN NS	HL	PO	NN TA	KS SK	IP FO	CR RC	CE EC	NO ON	IS
F	HL	IR -O	KP	LO OL	NO ON	NZ	UC -B	TA AT	LT TL	OI ES	IH HI	OC PS	NO ON	CE EC	RA AR
G	OS	DI	NO ON	EE EE	KR	LC	UB	RA AR	OS	DI	IP FO	DA	RC CR	OG	GR -E
H	OL -S	NO ON	CW	DI	LP	OI ES	LN T-	QX	DI	GL SA	RB N-	BQ	YF	SS L-	RA AR
J	VY	OI ES	GR -E	SL LS	XX										

Skeletons of additional plain text, such as the word OUR at A1, PRESENCE OF ENEMY at B1, PROBABLE at D1, ATTACK ON OUR INSTALLATIONS at D10, CARRIER at F14, and VESSELS at J1, may now clearly be seen. The complete recovery of the plain text follows, and the reconstruction diagram is completed and telescoped into the form shown in Fig. 64e. Since phenomena of keyword-mixed sequences are observed, the rows and columns of Fig. 64e are permuted to yield the original two-square matrix as shown in Fig. 64f.

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

Figure 64e.

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

Figure 64f.

h. The solution of vertical two-square systems follows analogous lines, with the necessary modifications of the reconstruction diagram in consonance with the difference in mechanics between horizontal and vertical two-square systems.

i. A few additional remarks concerning the test applied in subpars. d and e, above, are in order. First, the exceptionally high transparency value observed in this cryptogram is a direct result of the very favorable manner in which the keyword-mixed sequences in the two squares interact; in the foregoing cryptogram, 47 of the 125 digraphs present (approx. 38%) were inverse transparencies. It is also pointed out that, although some actual two-square cryptograms may be rejected by that portion of the test which was described in subpar. e, the other phase of the test (described in subpar. d)--by which one may determine whether a cryptogram is more probably a vertical two-square encipherment or more probably a horizontal two-square encipherment--is extremely sensitive and highly accurate. The foregoing statistical method is not merely valuable *per se* as an application of cryptomathematics in the analysis of two-square matrix systems, but is included as being illustrative of the general principles of special techniques that may be developed in the attack on any particular cryptosystem, the mechanics of which are known to the cryptanalyst. The field of actual operational cryptanalysis is replete with special methods of attack of this nature.

71. Analysis of Playfair cipher systems.--a. Of all digraphic cryptosystems employing small matrices, the one which has been most frequently encountered is the Playfair cipher. Certain variations of this cipher have been incorporated in several complex manual ciphers used in actual operational practice; because of this it is important that the student gain familiarity with the methods of solution of the classic Playfair system.

b. The first published solutions[32] for this cipher are quite similar basically and vary only in minor details. The earliest, that by Lieut. Mauborgne (later to become Chief Signal Officer of the U.S. Army), used straightforward principles of frequency to establish the values of three or four of the most frequent digraphs. Then, on the assumption that in most cases in which a keyword appears on the first and second rows the last five letters of the normal alphabet, VWXYZ, will rarely be disturbed in sequence and will occupy the last row of the square, he "juggles" the letters given by the values tentatively established from frequency considerations, placing them in various positions in the square, together with VWXYZ, to correspond to the plaintext-ciphertext relationships tentatively established. A later solution by Lieut. Frank Moorman, as described in Hitt's manual, assumes that in a Playfair cipher prepared by means of a square in which the key word occupies the first and second rows, if a digraphic frequency distribution is made, it will be found that the letters having the greatest combining power are very probably letters of the key. A still later solution, by Lieut. Commander Smith, is perhaps the most lucid and systematized of the three. He sets forth in definite language certain considerations which the other two writers certainly entertained but failed to indicate.

c. The following details have been summarized from Smith's solution:

(1) The Playfair cipher may be recognized by virtue of the fact that it always contains an even number of letters, and that when divided into groups of two letters each, no group contains a repetition of the same letter, as NN or EE. Repetitions of digraphs, trigraphs, and polygraphs will be evident in fairly long messages.

(2) Using the square[33] shown in Fig. 65, there are two general cases to be considered, as regards the results of encipherment:

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

Figure 65.

[32] Mauborgne, Lieut. J. O., U.S.A. An advanced problem in cryptography and its solution, Leavenworth, 1914.

Hitt, Captain Parker, U.S.A. Manual for the solution of military ciphers, Leavenworth, 1918.

Smith, Lieut. Commander W. W., U.S.N. In Cryptography by André Langie, translated by J.C.H. Macbeth, New York, 1922.

[33] The Playfair square accompanying Smith's solution is based upon the key word BANKRUPTCY, "to be distributed between the first and fourth lines of the square". This is a simple departure from the original Playfair scheme in which the letters of the key word are written from left to right and in consecutive lines from the top downward.

Case 1. Letters at opposite corners of a rectangle. The following illustrative relationships are found:

$$\overline{TH}_p=\overline{YF}_c$$
$$\overline{HT}_p=\overline{FY}_c$$
$$\overline{YF}_p=\overline{TH}_c$$
$$\overline{FY}_p=\overline{HT}_c$$

Reciprocity and reversibility.[34]

Case 2. Two letters in the same row or column. The following illustrative relationships are found:

$$\overline{AN}_p=\overline{NK}_c$$
$$\overline{NA}_p=\overline{KN}_c$$

But $\overline{NK}_p$ does not = $\overline{AN}_c$, nor does $\overline{KN}_p=\overline{NA}_c$.

Reversibility only.

(3) The foregoing gives rise to the following:

Rule I. (a) Regardless of the position of the letters in the square, if

1.2=3.4, then
2.1=4.3

This rule is of particular aid in selecting probable words in the solution of Playfair ciphers, as will be shown shortly.[35]

(b) If 1 and 2 form opposite corners of a rectangle, the following equations obtain:

1.2=3.4
2.1=4.3
3.4=1.2
4.3=2.1

[34] By way of explaining what is meant by reciprocity and by reversibility, in the case of digraphic systems, the following examples are given: $\overline{TH}_p=\overline{YF}_c$ and $\overline{YF}_p=\overline{TH}_c$ constitute a reciprocal relationship; $\overline{TH}_p=\overline{YF}_c$ and $\overline{HT}_p=\overline{FY}_c$ constitute a reversible relationship.

[35] In this connection, a list of frequently-encountered words and phrases which contain reversed digraphs (so-called "ABBA patterns") has been compiled and is included as Section E, "Digraphic idiomorphs: Playfair", in Appendix 3.

(4) A letter considered as occupying a position in a row can be combined with but four other letters in the same row; the same letter considered as occupying a position in a column can be combined with but four other letters in the same column. Thus, this letter can be combined with only 8 other letters all told, under Case 2, above. But the same letter considered as occupying a corner of a rectangle can be combined with 16 other letters, under Case 1, above. Smith derives from these facts the conclusion that "it would appear that Case 1 is twice as probable as Case 2". He continues thus (notation my own):

"Now in the square, note that:

$$\begin{array}{lcl} \overline{AN}_p = \overline{NK}_c & & \overline{EN}_p = \overline{FA}_c \\ \overline{GN}_p = \overline{FK}_c & & \overline{EM}_p = \overline{FL}_c \\ \overline{ON}_p = \overline{MK}_c & \text{also} & \overline{ET}_p = \overline{FP}_c \\ \overline{CN}_p = \overline{TK}_c & & \overline{EW}_p = \overline{FV}_c \\ \overline{XN}_p = \overline{WK}_c & & \overline{EF}_p = \overline{FG}_c \end{array}$$

"From this it is seen that of the 24 equations that can be formed when each letter of the square is employed either as the initial or final letter of the group, five will indicate a repetition of a corresponding letter of plain text.

"Hence, Rule II. After it has been determined, in the equation 1.2=3.4, that, say, $\overline{EN}_p = \overline{FA}_c$, there is a probability of one in five that any other group beginning with F_c indicates $\overline{E\theta}_p$, and that any group ending in A_c indicates $\overline{\theta N}_p$.[36]

"After such combinations as $\overline{ER}_p$, $\overline{OR}_p$, and $\overline{EN}_p$ have been assumed or determined, the above rule may be of use in discovering additional digraphs and partial words".

[36] There is an error in this reasoning. Take, for example, the 24 equations having F as an initial letter:

Case		Case		Case		Case	
1.	$FB_c = DN_p$	2.	FE=ED	2.	FT=NM	1.	FX=GW
2.	FD =EH	1.	FL=EM	2.	FW=NT	1.	FR=HN
1.	FI =DM	1.	FP=ET	1.	FK=GN	2.	FH=EG
1.	FU =DT	1.	FV=EW	2.	FG=EF	1.	FQ=HM
1.	FS =DW	2.	FN=NW	1.	FO=GM	1.	FY=HT
1.	FA =EN	2.	FM=NF	1.	FC=GT	1.	FZ=HW

Here, the initial letter F_c represents the following initial letters of plain-text digraphs:

$D\theta_p$, $E\theta_p$, $N\theta_p$, $G\theta_p$, and $H\theta_p$.

It is seen that F_c represents D_p, N_p, G_p, H_p 4 times each, and E_p, 8 times. Consequently, supposing that it has been determined that $FA_c = EN_p$, the probability that F_c will represent E_p is not 1 in 5 but 8 in 24, or 1 in 3; but supposing that it has been determined that $FW_c = NT_p$, the probability that F_c will represent N_p is 4 in 24 or 1 in 6. The difference in these probabilities is occasioned by the fact that the first instance, $FA_c = EN_p$ corresponds to a Case 1 encipherment, the second instance, $FW_c = NT_p$, to a Case 2 encipherment. But there is no way of knowing initially, and without other data, whether one is dealing with a Case 1 or Case 2 encipherment. Only as an approximation, therefore, may one say that the probability of F_c representing a given θ_p is 1 in 5. A probability of 1 in 5 is of almost trivial importance in this situation, since it represents such a "long shot" for success. The following rule might be preferable: If the equation 1.2=3.4 has been established, where all the letters represented

Rule III. In the equation $1.2=3.4$, 1 and 3 can never be identical, nor can 2 and 4 ever be identical. Thus, $\overline{\text{AN}}_p$ could not possibly be represented by $\overline{\text{AY}}_c$, nor could $\overline{\text{ER}}_p$ be represented by $\overline{\text{KR}}_c$. This rule is useful in elimination of certain possibilities when a specific message is being studied.

Rule IV. In the equation $1.2_p=3.4_c$, if 2 and 3 are identical, the letters are all in the same row or column, and in the relative order 1-2-4 from left to right or top to bottom, respectively. In the square shown, $\overline{\text{AN}}_p=\overline{\text{NK}}_c$ and the absolute order is ANK. The relative order 1-2-4 includes five absolute orders which are cyclic permutations of one another. Thus: ANK.., NK..A, K..AN, ..ANK, and .ANK..

Rule V. In the equation $1.2_p=3.4_c$, if 1 and 4 are identical, the letters are all in the same row or column, and in the relative order 2-4-3 from left to right or top to bottom. In the square shown, $\overline{\text{KN}}_p=\overline{\text{RK}}_c$ and the absolute order is NKR. The relative order 2-4-3 includes five absolute orders which are cyclic permutations of one another. Thus NKR.., KR..N, R..NK, ..NKR, and .NKR..

Rule VI. "Analyze the message for group recurrences. Select the groups of greatest recurrence and assume them to be high-frequency digraphs.[37] Substitute the assumed digraphs throughout the message, testing the assumptions in their relation to other groups of the cipher. The reconstruction of the square proceeds simultaneously with the solution of the message and aids in hastening the translation of the cipher".

by 1, 2, 3, and 4 are different, then there is a probability of 4/5 that a Case 1 encipherment is involved. Consequently, if at the same time another equation, $3.6=5.2$, has been established, where 2 and 3 represent the same letters as in the first equation, and 5 and 6 are different letters, also different from 2 and 3, there is a probability of 16/25 that the equation $1.6=5.4$ is valid; or if at the same time that the equation $1.2=3.4$ has been determined, the equation $1.6=5.4$ has also been established, then there is a probability of 16/25 that the equation $3.6=5.2$ is valid. (Check this by noting the following equations based upon Fig. 25*a*: $\overset{1\,2}{\text{CE}}=\overset{3\,4}{\text{PG}}$, $\overset{3\,6}{\text{PH}}=\overset{5\,2}{\text{YE}}$, $\overset{1\,6}{\text{CH}}=\overset{5\,4}{\text{YG}}$. Note the positions occupied in Fig. 25*a* by the letters involved.) Likewise, if the equations $1.2=3.4$ and $1.6=3.5$ have been simultaneously established, then there is a probability that the equation $2.5=4.6$ is valid; or if the equations $1.2=3.4$ and $2.5=4.6$ have been simultaneously established, then there is a probability that the equation $2.5=4.6$ is valid. (Check this by noting the following equations: $\overset{1\,2}{\text{CE}}=\overset{3\,4}{\text{PG}}$; $\overset{1\,6}{\text{CA}}=\overset{3\,5}{\text{PK}}$; $\overset{2\,5}{\text{EK}}=\overset{4\,6}{\text{GA}}$; note the positions occupied in Fig. 25*a* by the letters involved.) However, it must be added that these probabilities are based upon assumptions which fail to take into account any considerations whatever as to frequency of letters or specificity of composition of the matrix. For instance, suppose the 5 high-frequency letters E, T, R, I, N all happen to fall in the same row or column in the matrix; the number of Case 2 encipherments would be much greater than expectancy and the probability that the equation $1.2=3.4$ represents a Case 1 encipherment falls much below 4/5.

[37] A more accurate guide to the determination of the plaintext equivalents of high-frequency cipher digraphs would involve the consideration of the difference in frequency of a particular digraph and its reversal. Thus, an example of a high-frequency $\overline{\theta\theta}_p$ which is also high-frequency in its reversal, is $\overline{\text{RE}}_p$; an example of a high-frequency $\overline{\theta\theta}_p$ which is rarely found in its reversed form, is $\overline{\text{TH}}_p$.

d. (1) When solutions for the Playfair cipher system were first developed, based upon the fact that the letters were inserted in the cells in keyword-mixed order, cryptographers thought it desirable to place stumbling blocks in the path of such solution by departing from strict, keyword-mixed order. One of the simplest methods is illustrated in Fig. 65, wherein it will be noted that the last five letters of the keyword proper are inserted in the fourth row of the square instead of the second, where they would naturally fall. Another method involves inserting the letters within the cells from left to right and top downward but using a sequence that is derived from a columnar transposition instead of a keyword-mixed sequence. Thus, using the keyword BANKRUPTCY:

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

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

The Playfair square is as follows:

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

Figure 66a.

(2) Note the following three squares:

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

Figure 66b.

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

Figure 66c.

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

Figure 66d.

At first glance they all appear to be different, but closer examination shows them to be cyclic permutations of one another and of the square in Fig. 66a. They yield identical cryptographic equivalents in all cases. However, if an attempt be made to reconstruct the original key word, it would be much easier to do so from Fig. 66a than from any of the others, because in Fig. 66a the original keyword-mixed sequence has not been disturbed as much as in Figs. 66b, c, and d. In working with Playfair ciphers, the student should be on the lookout for such instances of cyclic permutation of the original Playfair square, for during the course of solution he will not know whether he is building up the original or an equivalent cyclic permutation of the original matrix; usually only after he has completely reconstructed the matrix will he be able to determine this point.

e. (1) The steps in the solution of a typical example of this cipher will now be illustrated. Let the message be 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 27 28 29 30
A.  V T Q E U   H I O F T   C  H  X  S  C   A  K  T  V  T   R  A  Z  E  V   T  A  G  A  E
B.  O X T Y M   H C R L Z   Z  T  Q  T  D   U  M  C  Y  C   X  C  T  G  M   T  Y  C  Z  U
C.  S N O P D   G X V X S   C  A  K  T  V   T  P  K  P  U   T  Z  P  T  W   Z  F  N  B  G
D.  P T R K X   I X B P R   Z  O  E  P  U   T  O  L  Z  E   K  T  T  C  S   N  H  C  Q  M
E.  V T R K M   W C F Z U   B  H  T  V  Y   A  B  G  I  P   R  Z  K  P  C   Q  F  N  L  V
F.  O X O T U   Z F A C X   X  C  P  Z  X   H  C  Y  N  O   T  Y  O  L  G   X  X  I  I  H
G.  T M S M X   C P T O T   C  X  O  T  T   C  Y  A  T  E   X  H  F  A  C   X  X  C  P  Z
H.  X H Y C T   X W L Z T   S  G  P  Z  T   V  Y  W  C  E   T  W  G  C  C   M  B  H  M  Q
J.  Y X Z P W   G R T I V   U  X  P  U  M   Q  R  K  M  W   C  X  T  M  R   S  W  G  H  B
K.  X C P T O   T C X O T   M  I  P  Y  D   N  F  G  K  I   T  C  O  L  X   U  E  T  P  X
L.  X F S R S   U Z T D B   H  O  Z  I  G   X  R  K  I  X   Z  P  P  V  Z   I  D  U  H  Q
M.  O T K T K   C C H X X
```

(2) Without going through the preliminary tests in detail, with which it will be assumed that the student is now familiar,[38] the conclusion is reached that the cryptogram is digraphic in nature, and a digraphic frequency distribution is made (Fig. 67).

[38] See par. 69c.

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

Figure 67

Since there are no double-letter groups (termed "doublets"), the conclusion is reached that a Playfair cipher is involved and the message is rewritten in digraphs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	VT	QE	UH	IO	FT	CH	XS	CA	KT	VT	RA	ZE	VT	AG	AE
B.	OX	TY	MH	CR	LZ	ZT	QT	DU	MC	YC	XC	TG	MT	YC	ZU
C.	SN	OP	DG	XV	XS	CA	KT	VT	PK	PU	TZ	PT	WZ	FN	BG
D.	PT	RK	XI	XB	PR	ZO	EP	UT	OL	ZE	KT	TC	SN	HC	QM
E.	VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
F.	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY	OL	GX	XI	IH
G.	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ
H.	XH	YC	TX	WL	ZT	SG	PZ	TV	YW	CE	TW	GC	CM	BH	MQ
J.	YX	ZP	WG	RT	IV	UX	PU	MQ	RK	MW	CX	TM	RS	WG	HB
K.	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET	PX
L.	XF	SR	SU	ZT	DB	HO	ZI	GX	RK	IX	ZP	PV	ZI	DU	HQ
M.	OT	KT	KC	CH	XX										

(3) The following three fairly lengthy repetitions are noted:

Lines									
F.	OT	UZ	FA	CX	XC	PZ	XH	CY	NO
G.	TE	XH	FA	$\overrightarrow{CX}$	$\overleftarrow{XC}$	PZ	XH	YC	TX
A.	FT	CH	XS	CA	KT	VT	RA	ZE	
C.	DG	XV	XS	CA	KT	VT	PK	PU	
G.	TM	SM	XC	PT	OT	CX	OT	TC	
K.	WG	HB	XC	PT	OT	CX	OT	MI	

The first long repetition, with the sequent reversed digraphs $\overline{CX}$ and $\overline{XC}$ immediately suggests the word BATTALION (see Section E, Appendix 3), split up into -B AT TA LI ON and the sequence containing this repetition in lines F and G becomes as follows:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
				B	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL.
				B	AT	TA	LI	ON			

(4) Because of the frequent use of numerals before the word BATTALION (as mentioned in Section B of Appendix 4) and because of the appearance of ON_p before this word in line G, the possibility suggests itself that the word before BATTALION in line G is either ONE or SECOND. The identical cipher digraph $\overline{FA}$ in both cases gives a hint that the word BATTALION in line F may also be preceded by a numeral; if ONE is correct

in line G, then THREE is possible in line F. On the other hand, if SECOND is correct in line G, then THIRD is possible in line F. Thus:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
1st hypothesis....	--	TH	RE	EB	AT	TA	LI	ON			
2nd hypothesis....	--	TH	IR	DB	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
1st hypothesis....	--	--	ON	EB	AT	TA	LI	ON			
2nd hypothesis....	-S	EC	ON	DB	AT	TA	LI	ON			

First, note that if either hypothesis is true, then $\overline{OT}_c=\overline{TH}_p$. The frequency distribution shows that $\overline{OT}$ occurs 6 times and is in fact the most frequent digraph in the message. Moreover, by Rule I of subparagraph b, if $\overline{OT}_c=\overline{TH}_p$ then $\overline{TO}_c=\overline{HT}_p$. Since $\overline{HT}_p$ is a very rare digraph in normal plain text, $\overline{TO}_c$ should either not occur at all in so short a message or else it should be very infrequent. The frequency distribution shows that it does not occur. Hence, there is nothing inconsistent with the supposition that the word in front of BATTALION in line F is THREE or THIRD, and there is some evidence that it is actually one or the other.

(5) But can evidence be found for the support of one hypothesis against the other? Let the frequency distribution be examined with a view to throwing light upon this point. If the first hypothesis is true, then $\overline{UZ}_c=\overline{RE}_p$, and, by Rule I, $\overline{ZU}_c=\overline{ER}_p$. The frequency distribution shows but one occurrence of $\overline{UZ}_c$ and but two occurrences of $\overline{ZU}_c$. These do not look very good for $\overline{RE}$ and $\overline{ER}$. On the other hand, if the second hypothesis is true, then $\overline{UZ}_c=\overline{IR}_p$ and, by Rule I, $\overline{ZU}_c=\overline{RI}_p$. The frequencies are much more favorable in this case. Is there anything inconsistent with the assumption, on the basis of the second hypothesis, that $\overline{TE}_c=\overline{EC}_p$? The frequency distribution shows no inconsistency, for $\overline{TE}_c$ occurs once and $\overline{ET}_c$($=\overline{CE}_p$, by Rule I) occurs once. As regards whether $\overline{FA}_c=\overline{EB}_p$ or $=\overline{DB}_p$, both hypotheses are tenable; possibly the second hypothesis is a shade better than the first, on the following reasoning: By Rule I, if $\overline{FA}_c=\overline{EB}_p$ then $\overline{AF}_c=\overline{BE}_p$, or if $\overline{FA}_c=\overline{DB}_p$ then $\overline{AF}_c=\overline{BD}_p$. The fact that no $\overline{AF}_c$ occurs, whereas at least one $\overline{BE}_p$ may be expected in this message, inclines one to the second hypothesis, since $\overline{BD}_p$ is very rare.

(6) Let the 2nd hypothesis be assumed to be correct. The additional values are tentatively inserted in the text, and in lines G and K two interesting repetitions are noted:

Line G	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ	XH
			TA		TH	AT	TH		-S	EC	ON	DB	AT	TA	LI	ON
Line K	WG	HB	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET
			TA		TH	AT	TH									

This certainly looks like STATE THAT THE . . ., which would make $\overline{TE}_p=\overline{PT}_c$. Furthermore, in line G the sequence STATETHATTHE..SECONDBATTALION can hardly be anything else than STATE THAT THEIR SECOND BATTALION, which would make $\overline{TC}_c=\overline{EI}_p$ and $\overline{YA}_c=\overline{RS}_p$. Also $\overline{SM}_c=\overline{-S}_p$.

(7) It is perhaps high time that the whole list of tentative equivalent values be studied in relation to their consistency with the positions of letters in the Playfair square; moreover, by so doing, additional values may be obtained in the process. The complete list of values is as follows:

Assumed values	Derived by Rule I
$\overline{AT}_p=\overline{CX}_c$	$\overline{TA}_p=\overline{XC}_c$
$\overline{LI}_p=\overline{PZ}_c$	$\overline{IL}_p=\overline{ZP}_c$
$\overline{ON}_p=\overline{XH}_c$	$\overline{NO}_p=\overline{HX}_c$
$\overline{TH}_p=\overline{OT}_c$	$\overline{HT}_p=\overline{TO}_c$
$\overline{IR}_p=\overline{UZ}_c$	$\overline{RI}_p=\overline{ZU}_c$
$\overline{DB}_p=\overline{FA}_c$	$\overline{BD}_p=\overline{AF}_c$
$\overline{EC}_p=\overline{TE}_c$	$\overline{CE}_p=\overline{ET}_c$
$\overline{TE}_p=\overline{PT}_c$	$\overline{ET}_p=\overline{TP}_c$
$\overline{EI}_p=\overline{TC}_c$	$\overline{IE}_p=\overline{CT}_c$
$\overline{RS}_p=\overline{YA}_c$	$\overline{SR}_p=\overline{AY}_c$
$\overline{-S}_p=\overline{SM}_c$	$\overline{S-}_p=\overline{MS}_c$

(8) By Rule V, the equation $\overline{TH}_p=\overline{OT}_c$ means that H, O, and T are all in the same row or column and in the absolute order HTO; similarly, C, E, and T are in the same row or column and in the absolute order CET. Further, E, P, and T are in the same row and column, and their absolute order is ETP. That is, these sequences must occur someplace in the square, in either rows or columns, taking into consideration of course the probability of cyclic displacements of these sequences within the square:

(a) H T O (b) C E T (c) E T P

(9) Noting the common letters E and T in the second and third sequences, these two sequences may be combined into one sequence of four letters, viz., C E T P. Since only one position remains to be filled in this row (or column) of the square, and noting in the list of equivalents that $\overline{EI}_p=\overline{TC}_c$, it is obvious that the letter I belongs to the C E T P sequence; the complete sequence is therefore C E T P I.

(10) Since the sequence HTO has a common letter (T) with the sequence CETPI, it follows that if the HTO sequence occupies a row, then the CETPI sequence must occupy a column; or, if the HTO sequence occupies a column, then the CETPI sequence must occupy a row; and they may be combined by means of their common letter, T, viz.:

		H		
C	E	T	P	I
		O		

or

	C	
	E	
H	T	O
	P	
	I	

The proof of whether the CETPI sequence, for example, properly belongs as a row or a column of the Playfair square lies in the establishment of a rectangular relationship, instead of the linear relationships constructed thus far.

(11) We note that, from the assumptions in subpar. d(6), $\overline{AT}_p=\overline{CX}_c$ and $\overline{ON}_p=\overline{XH}_c$. The relationship $\overline{ON}_p=\overline{XH}_c$ might be either a rectangular one, such as

```
O   X
 \ /
  X
 / \
H   N
```

or it might be linear, viz., HTOXN or

```
H
T
O
X
N
```

Since however $\overline{AT}_p=\overline{CX}_c$ must be a rectangular relationship, then only the configuration

```
 C   A
 E
H T O X N
 P
 I
```

will be valid, since the alternative form

```
   H
C E T P I
   O
A  X
   N
```

will not satisfy the equation $\overline{AT}_p=\overline{CX}_c$.

(12) The fragmentary Playfair square[39] has been established, in one of its 25 possible cyclic permutations, as

```
 C   A
 E
H T O X N
 P
 I
```

Scanning the list of plain-cipher equivalents given in subpar. d(7) in order to insert possible additional letters, none is found. But seeing that several high-frequency letters have already been inserted in the matrix, perhaps reference to the cryptogram itself in connection with values derived from these inserted letters may yield further clues. For example, the vowels A, E, I, and O are all in position, as are the very frequent consonants N and T. The following combinations may be studied:

$\overline{AN}_p=\overline{\theta X}_c$	$\overline{AT}_p=\overline{CX}_c$	$\overline{NA}_p=\overline{X\theta}_c$	$\overline{TA}_p=\overline{XC}_c$
$\overline{EN}_p=\overline{\theta T}_c$	$\overline{ET}_p=\overline{TP}_c$	$\overline{NE}_p=\overline{T\theta}_c$	$\overline{TE}_p=\overline{PT}_c$
$\overline{IN}_p=\overline{\theta T}_c$	$\overline{IT}_p=\overline{CP}_c$	$\overline{NI}_p=\overline{T\theta}_c$	$\overline{TI}_p=\overline{PC}_c$
$\overline{ON}_p=\overline{XH}_c$	$\overline{OT}_p=\overline{XO}_c$	$\overline{NO}_p=\overline{HX}_c$	$\overline{TO}_p=\overline{OX}_c$

[39] In actual practice, it is more usual to start with a much larger diagram than a simple 5x5 square; as relationships develop, the diagram is gradually condensed, until finally a 5x5 square emerges. This procedure is quite similar to that employed in the reconstruction diagrams for two-square matrices.

$\overline{AT}_p(=\overline{CX}_c)$, $\overline{TA}_p(=\overline{XC}_c)$, $\overline{ON}_p(=\overline{XH}_c)$, $\overline{TE}_p(=\overline{PT}_c)$ and $\overline{ET}_p(=\overline{TP}_c)$ have already been inserted in the text. Of the others, only $\overline{OX}_c(=\overline{TO}_p)$ occurs two times, and this value can be at once inserted in the text. But can the equivalents of $\overline{AN}$, $\overline{EN}$, or $\overline{IN}$ be found from frequency considerations? Take $\overline{EN}_p$, for example; it is represented by $\overline{\theta T}_c$. What combination of $\overline{\theta T}$ is most likely to represent $\overline{EN}_p$ among the following candidates:

$\overline{KT}_c$ (4 times); by Rule I, $\overline{NE}_p$ would = $\overline{TK}_c$ (no occurrences)

$\overline{VT}_c$ (5 times); by Rule I, $\overline{NE}_p$ would = $\overline{TV}_c$ (2 times)

$\overline{ZT}_c$ (3 times); by Rule I, $\overline{NE}_p$ would = $\overline{TZ}_c$ (1 time)

$\overline{VT}_c$ certainly looks good: it begins the message, suggesting the word ENEMY, and the sequence $PZTV_c$, in line H, would become the plaintext sequence LINE. Let this be assumed to be correct, and let the word ENEMY also be assumed to be correct. Then $\overline{EM}_p=\overline{QE}_c$ and the partial square then becomes as shown herewith:

		P		
		I		
		C		A
V	M	E	Q	
N	H	T	O	X

Figure 68a.

(13) In line E is seen the following sequence:

Line E	VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
	EN				RI		NE	RS		PT			-E		

The plaintext sequence ...RI..NERS..PT... suggests PRISONERS CAPTURED, as follows:

MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP
	P	RI	SO	NE	RS	CA	PT	UR	ED

This gives the following new values: $\overline{\theta P}_p=\overline{CF}_c$; $\overline{SO}_p=\overline{BH}_c$; $\overline{CA}_p=\overline{BG}_c$; $\overline{UR}_p=\overline{RZ}_c$; and $\overline{ED}_p=\overline{KP}_c$. The letters B and G can be placed in position in the partial square at once, since the positions of C and A are already known. The insertion of the letter B immediately permits the placement of the letter S, from the equation $\overline{SO}_p=\overline{BH}_c$. Of the remaining equations only $\overline{ED}_p=\overline{KP}_c$ can be used. Since E and P are fixed and are in the same column, D and K must be in the same column, and moreover the K must be in the same row as E. There is only one possible position for K, *viz.*, immediately after Q. This automatically fixes the position of D. The square is now as shown herewith:

		P		D
		I		
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

Figure 68*b*.

(14) A review of all equations, including the very first ones established, gives the following which may now be used: $\overline{DB}_p=\overline{FA}_c$; $\overline{RS}_p=\overline{YA}_c$. The first permits the immediate placement of F; the second, by elimination of possible positions, permits the placement of both R and Y. The sq e is now as shown herewith:

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

Figure 68*c*.

Once more a review is made of all remaining unused equations. $\overline{LI}_p=\overline{PZ}_c$ now permits the placement of L and Z. $\overline{IR}_p=\overline{UZ}_c$ now permits the placement of U, which is confirmed by the equation $\overline{UR}_p=\overline{RZ}_c$ from the word CAPTURED. There is then only one cell vacant, and it must be occupied by the only letter left unplaced, *viz.*, W. Thus the whole square has been reconstructed, and the message can now be deciphered.

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

Figure 68*d*.

f. Reconstruction of the square in Playfair ciphers is normally carried on concurrently with the synthesis of the plain text, once a few correct assumptions have been made. Now, having just reconstructed the square as shown in Fig. 68*d*, the question to be answered is whether this square is identical with the original enciphering matrix or whether it is a cyclic permutation of the original square (which may have contained, say, a transposition-mixed sequence). Even though the cryptogram in subpar. 71*e* has been solved, this point is still of interest.

(1) The square that is derived may not necessarily be the original enciphering square; more than likely it will be one of the 24 possible cyclic permutations of the original square. If the Playfair square consisted of a keyword-mixed sequence, a permutation of the square will cause no difficulty in recovering the original matrix and hence the key word. For example, if the square derived in some other instance is

```
Q T L N O then the square P Y R A M is easily recovered because of the
X Z U V W                 I D S B C
A M P Y R                 E F G H K
B C I D S                 L N O Q T
H K E F G                 U V W X Z
```

tell-tale letters UVWXZ occurring in a row of the derivative square. But when the Playfair square consists of a transposition-mixed sequence, then a different procedure must be adopted.

(2) As an example, let us take the transposition matrix

```
5 8 6 1 4 3 2 7 from which A F T D K is the original square.  Using the
P Y R A M I D S            W I H V M
B C E F G H K L            G U P B N
N O Q T U V W X            Z R E Q S
Z                          L X Y C O
```

methods illustrated in par. 51g, scanning successive rows of the square will disclose sequences of letters which could have appeared as columns in the transposition matrix. For example, discovery of the columns

```
I|D|S
H|K|L
V|W|X
```

will afford rapid recovery of the key word. But if instead of the original square we had one of its permutations such as

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

of the "columns", e.g.,

```
F|V|O|L|Q, of the tentative transposition matrix
T|M|L|X|S
V|W|X|Y|Z
```

(assuming that some or all of the letters V, W, X, Y, Z are in the last row of the transposition matrix) will be without significance; therefore the procedure above is inapplicable without a slight modification.

(3) Since it will be noted that a permutation of the rows will not affect the procedure of keyword recovery, then we construct a 9x5 rectangle

```
Q S Z R E Q S Z R which contains the five squares which result
C O L X Y C O L X
D K A F T D K A F
V M W I H V M W I
B N G U P B N G U
```

simply from successive permutations of the columns. A 5x5 cut-out square will be found convenient in testing each permutation in turn. Affirmative results will be obtained when the correct permutation is reached,

which in this case is the third square in the rectangle, namely,
Z R E Q S. After recovery of the key word from this permuted square it
L X Y C O
A F T D K
W I H V M
G U P B N

is probable then that the original enciphering square must have been

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

(4) In the case of the square recovered in Fig. 68d, it is found that, following the procedure outlined in subpars. (1), (2), and (3) above he key word is based on COMPANY, recoverable from the following diagram:

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

The original square must have been this:

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

Figure 68e.

g. Continued practice in the solution of Playfair ciphers will make the student quite expert in the matter and will enable him to solve shorter and shorter messages.[40] Also, with practice it will become a matter of indifference to him as to whether the letters are inserted in the square with any sort of regularity, such as simple keyword-mixed order, transposition-mixed order, or in a purely random order.

h. It may perhaps seem to the student that the foregoing steps are somewhat too artificial, a bit too "cut and dried" in their accuracy to portray the process of analysis as it is applied in practice. For example, the critical student may well object to some of the assumptions and the reasoning in subpar. e(5), above, in which the words THREE and

[40] The author once had a student who "specialized" in Playfair ciphers and became so adept that he could solve messages containing as few as 50-60 letters within 30 minutes.

ONE (1st hypothesis) were rejected in favor of the words THIRD and SECOND (2nd hypothesis). This rested largely upon the rejection of $\overline{RE}_p$ and $\overline{ER}_p$ as the equivalents of $\overline{UZ}_c$ and $\overline{ZU}_c$, and the adoption of $\overline{IR}_p$ and $\overline{RI}_p$ as their equivalents. Indeed, if the student will examine the final message with a critical eye, he will find that while the bit of reasoning in step (5) is perfectly logical, the assumption upon which it is based is in fact wrong; for it happens that in this case $\overline{ER}_p$ occurs only once and $\overline{RE}_p$ does not occur at all. Consequently, although most of the reasoning which led to the rejection of the first hypothesis and the adoption of the second was logical, it was in fact based upon erroneous assumption. In other words, despite the fact that the assumption was incorrect, a correct deduction was made. The student should take note that in cryptanalysis situations of this sort are not at all unusual. Indeed they are to be expected, and a few words of explanation at this point may be useful.

i. Cryptanalysis is a science in which deduction, based upon observational data, plays a very large role. But it is also true that in this science most of the deductions usually rest upon assumptions. It is most often the case that the cryptanalyst is forced to make his assumptions based upon a quite limited amount of text. It cannot be expected that assumptions based upon statistical generalizations will always hold true when applied to data comparatively very much smaller in quantity than the total data used to derive the generalized rules. Consequently, as regards assumptions made in specific messages, most of the time they will be correct, but occasionally they will be incorrect.[41] In cryptanalysis it is often found that among the correct deductions there will be cases in which subsequently discovered facts do not bear out the assumptions on which the deduction was based. Indeed, it is sometimes true that if the facts had been known before the deduction was made, this knowledge would have prevented making the correct deduction. For example, suppose the cryptanalyst had somehow or other divined that the message under consideration contained no RE, only one ER, one IR, and two RI's (as is actually the case). He would certainly not have been able to choose between the words THREE and ONE (1st hypothesis) as against THIRD and SECOND (2d hypothesis). But because he assumes that there should be more $\overline{ER}_p$'s and $\overline{RE}_p$'s than $\overline{IR}_p$'s and $\overline{RI}_p$'s in the message, he deduces that $\overline{UZ}_c$ cannot be $\overline{RE}_p$, rejects the first hypothesis and takes the second. It later turns out, after the problem has been solved, that the deduction was correct, although the assumption on which it was based (expectation of more frequent appearance of $\overline{RE}_p$ and $\overline{ER}_p$) was, in fact, not true in this particular case. The cryptanalyst can only hope that the number of times when his deductions are correct, even though based upon assumptions which later turn out to be erroneous, will abundantly exceed the number of times when his deductions are wrong, even though based upon assumptions which later prove to be correct. If he is lucky,

[41] See footnote 18 on page 52.

the making of an assumption which is really not true will make no difference in the end and will not delay solution; but if he is specially favored with luck, it may actually help him solve the message--as was the case in this particular example.

j. Another comment of a general nature may be made in connection with this specific example. The student may ask what would have been the procedure in this case if the message had not contained such a tell-tale repetition as the word BATTALION, which formed the point of departure for the solution, or, as it is often said, permitted an "entering wedge" to be driven into the message. The answer to his query is that if the word BATTALION had not been repeated, there would probably have been some other repetition which would have permitted the same sort of attack. If the student is looking for cut and dried, straightforward, unvarying methods of attack, he should remember that cryptanalysis, while considered a branch of mathematics by some, is not a science which has many "general solutions" such as are found and expected in mathematics proper. It is inherent in the very nature of cryptanalytics that, as a rule, only general principles can be established; their practical application must take advantage of peculiarities and particular situations which are noted in specific messages. This is especially true in a text on the subject. The illustration of a general principle requires a specific example, and the latter must of necessity manifest characteristics which make it different from any other example. The word BATTALION was not purposely repeated in this example in order to make the demonstration of solution easy; "it just happened that way". In another example, some other entering wedge would have been found. The student can be expected to learn only the general principles which will enable him to take advantage of the specific characteristics manifested in specific cases. Here it is desired to illustrate the general principles of solving Playfair ciphers and to point out the fact that entering wedges must and can be found. The specific nature of the entering wedge varies with specific examples.

72. Analysis of polygraphic systems involving large tables.--a. The analysis of systems incorporating large digraphic tables is accomplished by entering, within the appropriate cells of a 26x26 chart, data corresponding to the plain-cipher relationships of assumed cribs on 26x26 charts, and examining the charts for evidences of symmetry or systematic construction in their compilation. The initial plaintext entries may, in the absence of cribs, be made on the basis of digraphic frequency considerations, aided by idiomorphisms and repetitions.

b. In pseudo-digraphic systems, such as those incorporating tables similar to Figs. 47a and b, and 48, the identification of the monoalphabetically-enciphered component of cipher digraphs will greatly accelerate plaintext entries, since advantage may be taken of this monoalphabeticity. Tables with a feature of reciprocity, such as the example in Fig. 50, may be exploited on the basis of this weakness, even if the reciprocal pairs are assigned at random. Tables such as that in Fig. 49 and the one for trinome digraphic encipherment shown in Fig. 51 may also be exploited with facility, once enough plain text has been correctly assumed and inserted to disclose their systematic construction. A word of warning is inserted here against making incautious assumptions concerning the exact internal composition of tables such as that in Fig. 49, since their unusual construction could easily mislead the analyst who jumps to premature conclusions. In the case of a table such as Fig. 51 wherein the trinomes have been inscribed in straight horizontals (or for that matter, any other known inscription), if the dimensions of the table have been correctly assumed the simplest solution involves a reduction to two alphabets, reflecting the sequences of letters for the side and top of the matrix; this solution closely parallels that of the numerical four-square system described in subpar. 69e.

c. Because the foregoing principles are rather straightforward, it is not considered necessary to illustrate their application with examples. Of course, when digraphic tables of random construction have been used, no refinements in solution are possible. However, the recording of as few as 225 different plaintext digraphs and their ciphertext equivalents will theoretically enable the automatic decryption of approximately 92% of the cipher digraphs of messages, and the recording of 335 plaintext-ciphertext values will enable the automatic decryption of 98% of the cipher digraphs; thus almost every message may be read in its entirety without recourse to further assumptions. Actually, it should be pointed out that having only 122 matched plaintext-ciphertext equivalencies will theoretically enable the decryption of 75% of the cipher digraphs, and enough skeletons of plain text may then be manifest to permit the decryption of the complete message texts.

d. It might be well to point out in connection with large digraphic tables that there exist literal types which give rise to monoalphabetic distributions for both the initial letters and final letters of pairs. Such a table is illustrated in Fig. 69 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
A	HQ	YQ	DQ	RQ	AQ	UQ	LQ	IQ	CQ	BQ	EQ	FQ	GQ	JQ	KQ	MQ	NQ	OQ	PQ	QQ	SQ	TQ	VQ	WQ	XQ	ZQ
B	HU	YU	DU	RU	AU	UU	LU	IU	CU	BU	EU	FU	GU	JU	KU	MU	NU	OU	PU	QU	SU	TU	VU	WU	XU	ZU
C	HE	YE	DE	RE	AE	UE	LE	IE	CE	BE	EE	FE	GE	JE	KE	ME	NE	OE	PE	QE	SE	TE	VE	WE	XE	ZE
D	HS	YS	DS	RS	AS	US	LS	IS	CS	BS	ES	FS	GS	JS	KS	MS	NS	OS	PS	QS	SS	TS	VS	WS	XS	ZS
E	HT	YT	DT	RT	AT	UT	LT	IT	CT	BT	ET	FT	GT	JT	KT	MT	NT	OT	PT	QT	ST	TT	VT	WT	XT	ZT
F	HI	YI	DI	RI	AI	UI	LI	II	CI	BI	EI	FI	GI	JI	KI	MI	NI	OI	PI	QI	SI	TI	VI	WI	XI	ZI
G	HO	YO	DO	RO	AO	UO	LO	IO	CO	BO	EO	FO	GO	JO	KO	MO	NO	OO	PO	QO	SO	TO	VO	WO	XO	ZO
H	HN	YN	DN	RN	AN	UN	LN	IN	CN	BN	EN	FN	GN	JN	KN	MN	NN	ON	PN	QN	SN	TN	VN	WN	XN	ZN
I	HA	YA	DA	RA	AA	UA	LA	IA	CA	BA	EA	FA	GA	JA	KA	MA	NA	OA	PA	QA	SA	TA	VA	WA	XA	ZA
J	HB	YB	DB	RB	AB	UB	LB	IB	CB	BB	EB	FB	GB	JB	KB	MB	NB	OB	PB	QB	SB	TB	VB	WB	XB	ZB
K	HL	YL	DL	RL	AL	UL	LL	IL	CL	BL	EL	FL	GL	JL	KL	ML	NL	OL	PL	QL	SL	TL	VL	WL	XL	ZL
L	HY	YY	DY	RY	AY	UY	LY	IY	CY	BY	EY	FY	GY	JY	KY	MY	NY	OY	PY	QY	SY	TY	VY	WY	XY	ZY
M	HC	YC	DC	RC	AC	UC	LC	IC	CC	BC	EC	FC	GC	JC	KC	MC	NC	OC	PC	QC	SC	TC	VC	WC	XC	ZC
N	HD	YD	DD	RD	AD	UD	LD	ID	CD	BD	ED	FD	GD	JD	KD	MD	ND	OD	PD	QD	SD	TD	VD	WD	XD	ZD
O	HF	YF	DF	RF	AF	UF	LF	IF	CF	BF	EF	FF	GF	JF	KF	MF	NF	OF	PF	QF	SF	TF	VF	WF	XF	ZF
P	HG	YG	DG	RG	AG	UG	LG	IG	CG	BG	EG	FG	GG	JG	KG	MG	NG	OG	PG	QG	SG	TG	VG	WG	XG	ZG
Q	HH	YH	DH	RH	AH	UH	LH	IH	CH	BH	EH	FH	GH	JH	KH	MH	NH	OH	PH	QH	SH	TH	VH	WH	XH	ZH
R	HJ	YJ	DJ	RJ	AJ	UJ	LJ	IJ	CJ	BJ	EJ	FJ	GJ	JJ	KJ	MJ	NJ	OJ	PJ	QJ	SJ	TJ	VJ	WJ	XJ	ZJ
S	HK	YK	DK	RK	AK	UK	LK	IK	CK	BK	EK	FK	GK	JK	KK	MK	NK	OK	PK	QK	SK	TK	VK	WK	XK	ZK
T	HM	YM	DM	RM	AM	UM	LM	IM	CM	BM	EM	FM	GM	JM	KM	MM	NM	OM	PM	QM	SM	TM	VM	WM	XM	ZM
U	HP	YP	DP	RP	AP	UP	LP	IP	CP	BP	EP	FP	GP	JP	KP	MP	NP	OP	PP	QP	SP	TP	VP	WP	XP	ZP
V	HR	YR	DR	RR	AR	UR	LR	IR	CR	BR	ER	FR	GR	JR	KR	MR	NR	OR	PR	QR	SR	TR	VR	WR	XR	ZR
W	HV	YV	DV	RV	AV	UV	LV	IV	CV	BV	EV	FV	GV	JV	KV	MV	NV	OV	PV	QV	SV	TV	VV	WV	XV	ZV
X	HW	YW	DW	RW	AW	UW	LW	IW	CW	BW	EW	FW	GW	JW	KW	MW	NW	OW	PW	QW	SW	TW	VW	WW	XW	ZW
Y	HX	YX	DX	RX	AX	UX	LX	IX	CX	BX	EX	FX	GX	JX	KX	MX	NX	OX	PX	QX	SX	TX	VX	WX	XX	ZX
Z	HZ	YZ	DZ	RZ	AZ	UZ	LZ	IZ	CZ	BZ	EZ	FZ	GZ	JZ	KZ	MZ	NZ	OZ	PZ	QZ	SZ	TZ	VZ	WZ	XZ	ZZ

Figure 69.

In effect, encipherment by means of such a system yields the equivalent of a two-alphabet cipher, with a transposition within each of the pairs of letters. The cipher text produced by such a system may be characterized by a large number of repetitions which begin with the initial letter of digraphs and end on the final letter of digraphs and which are preceded by digraphs having repeated <u>initial</u> letters or which are followed by digraphs having repeated <u>final</u> letters; for example, ciphertext passages of the following type might often arise: <u>SF BD GB HK</u> and <u>SQ BD GB WK</u> (wherein the repeated plain text is actually represented by SDBBGK, affected by the transposition). This system is included here as being illustrative of many simple systems which are capable of leading the student very much astray; in this instance, if one were unaware of the transposition feature involved and were to attempt what appears to be the simple task of fitting plain text into the two monoalphabetic portions on the basis of single-letter frequency considerations, he could spend a great deal of time without success--probably without any idea of what was causing his difficulties.

<u>e</u>. A pseudo-trigraphic cipher involving a table such as that in Fig. 52 may be readily recognized as such, since two letters of each trigraph enciphered by means of such a table are treated monoalphabetically. If three separate unilateral frequency distributions are made--one for each of the three letters of the cipher trigraphs--two of the distributions should be monoalphabetic. Then, exploiting the monoalphabeticity

(i.e., the *positional* monoalphabeticity) thus disclosed in the cipher text, plain text can be fitted to the cipher on the basis of single-letter frequency considerations; in addition, advantage may be taken of *partial* idiomorphisms, if these idiomorphisms involve the particular positions of the trigraphs which have been treated monoalphabetically.

f. Fortunately, it is unlikely that trigraphic systems other than the foregoing pseudo-trigraphic type will be encountered, because they are difficult to manipulate without extensive tables or complicated rules for encryption.[41] The subject can be passed over with the simple statement that their analysis requires much text to permit of solution by the frequency method,--and blood, sweat, and tears.[42].

73. Further remarks on polygraphic substitution systems.--a. In the treatment of the cryptography of the various digraphic systems in this Section, the rules for encryption and decryption which have been illustrated are the "standard" rules (i.e., the rules extant in cryptologic literature, or the rules most commonly encountered in operational practice). Needless to say, however, there is no cryptologic counterpart of the Geneva Convention making these rules sacrosanct, nor forbidding the use of other rules for enciphering and deciphering.

b. In two-square systems and Playfair systems there are possible (and, in fact, there have been encountered in operational practice) modifications of the usual enciphering and deciphering rules which, if not suspected, may pose difficulties in the identification of such systems and in their cryptanalysis. For example, in a vertical two-square system, when two plaintext letters fall in the same column, their cipher equivalents might be taken as the letters immediately to the right of or immediately below these plaintext letters. Similarly, in a horizontal two-square system, if two plaintext letters are in the same row, their cipher equivalents might be taken as those immediately below, or to the right of these letters. In Playfair cipher systems, two plaintext letters in the same row might be represented by the letters immediately below; two plaintext letters in the same column might be represented by the letters immediately to the right; a plaintext doublet might be represented by a ciphertext *doublet* formed by doubling the letter immediately to the right, or below, or diagonally to the right and below, thus removing one of the identifying ciphertext characteristics of the *normal* Playfair system. In *one* case encountered, instead of the normal Playfair linear relationship $\overline{AB}_p=\overline{BC}_c$, the rule was changed to $\overline{AB}_p=\overline{CB}_c$ (thus allowing a

[41] However, see in this connection Appendix 8, "Lester S. Hill algebraic encipherment", which gives a mathematical treatment of *true* polygraphic encipherment for polygraphs of *any* size. (See also subpar. 73*h*).

[42] If a trigraphic system is encountered in operational cryptanalysis, special solutions would be made possible by the application of cribs, the aid furnished by isologs (not only in the same system, but also *between* systems), etc.

letter to "represent itself"--an "impossibility" in Playfair encipherment); even this simple modification caused difficulties in cryptanalysis because variant rules for encryption had not been considered.

c. The placing of cribs in small-matrix digraphic systems may be guided by the cryptographic peculiarities of these systems, when the general system is known to, or suspected by the cryptanalyst; conversely, the placing of a known crib may assist in the determination of the type of cryptosystem, or in the rejection of other types of systems. For example, cribs may be placed in Playfair ciphers on the basis of the "non-crashing" feature of the normal Playfair; that is, on the basis that in the equation 1.2=3.4 neither 1 and 3 nor 2 and 4 can be identical. In horizontal two-square systems, if $\alpha\beta_c = \_\alpha_p$, then $\alpha\beta_c$ must equal $\beta\alpha_p$; and if $\alpha\beta_c = \beta\_\,_p$, then $\alpha\beta_c$ must equal $\beta\alpha_p$. If, by placing a known crib in a cryptogram, evidence of non-reciprocity is disclosed (e.g., if $\overline{AB}_p = CD_c$, but $\overline{CD}_p = XY_c$), the cryptogram may be assumed to be other than a vertical two-square cipher, since vertical two-square encipherment yields complete reciprocity. In either type of two-square system, if one of the two squares is known (for example, a vertical two-square might be employed in which the upper square is always a normal alphabet), the placement of cribs is materially facilitated.

d. The ϕ test performed separately on the initial letter and final letter of ciphertext pairs from cryptograms produced by small-matrix digraphic systems will give results neither close to that expected for plain text, nor close to that for random text. The reason for the comparative "roughness" or pronounced differences among the relative frequencies in these distributions, as contrasted with the "smoothness" expected of random, is that small-matrix digraphic systems are only partially digraphic in nature and that the encryption involves characteristics similar to those of monoalphabetic substitution with variants. This roughness of the uniliteral frequency distributions for the prefixes and suffixes, and, for that matter, for the over-all cipher text, reflects the partially digraphic nature of the encipherment.

e. If the cipher letters V, W, X, Y, and Z are of very low frequency in the over-all uniliteral frequency distribution of a digraphic cryptogram or set of cryptograms, this may be taken as evidence that the cryptosystem is a small-matrix digraphic system employing keyword-mixed sequences in the matrix or matrices. Furthermore, in small-matrix systems involving keyword-mixed squares, if θ_c^1 of $\overline{\theta\theta}_c$ is one of the letters VWXYZ, the θ_p^1 of the corresponding $\overline{\theta\theta}_p$ is likely to be one of these same letters. Similarly, if θ_c^2 is one of the letters VWXYZ, then θ_p^2 of the corresponding $\overline{\theta\theta}_p$ is likely to be one of these letters.

f. In trinome-digraphic systems employing large tables, the trinomes may run from 001 to 676, as in Fig. 51, or any consecutive set of 676 trinomes in the scale of 1000 possible trinomes may be used. For that matter, the entire span of trinomes 000-999 might be used in such a table, with occasional gaps, to hide the limitations of this system. As another means of disguising the limitation of 676 trinomes in such a system, three of the initial digits of the trinomes might have one variant each--thus no limitation would exist in the first position of trinomes. The 001, or other starting point in the cyclic scale, need not be at the upper left-hand corner of the table. The 676 trinomes in such tables may be inscribed in straight horizontals (i.e., in the normal manner of writing) as in Fig. 51, or they might be inscribed according to some other route; they probably would not be inscribed in a random manner because clumsy "deciphering tables" would then be necessary. It is also possible that the trinomes in a trinome-digraphic system might be converted into tetranomes by the addition of a sum-check (to assist in error-correction).

g. The cryptanalysis of tetranome-trigraphic systems with matrices similar to that illustrated in Fig. 59 involves a modification of the technique used in solving inverse four-square systems. If the plain-component and cipher-component sections of the large square have been inscribed according to the normal manner of writing (or any other manner, if known), the first two elements of the trigraphs may be reduced to a pair of cipher alphabets, and these two monoalphabetic substitutions may be solved as indicated in subpar. 69e. The applicability of inverse four-square solution principles to this tetranome-trigraphic system of course rests on the fact that the ciphertext sections are known or assumed to contain the dinomes 00-99 in numerical order, inscribed in the normal manner of writing; the conversion of the first two elements of the trigraphs depends upon the knowledge of the manner of inscription of the letters of the plain component sections, in order that the four occurrences of the initial letters and the four occurrences of the final letters may be correctly combined into two monoalphabetic distributions. Of course, if the composition of the small square (for the third element of trigraphs) is known, the third letter of trigraphs may be automatically deciphered. If the composition of the small square is not known, a consideration of the frequencies of the converted dinomes for the small square (i.e., the coordinates of the square to indicate the third member of trigraphs) may be used to obtain an entering wedge into this third monoalphabetic substitution.

h. There are but a very limited number of known cipher mechanisms which employ the polygraphic encipherment principle in any form. U.S. Patent No. 1515680 issued to A. Henkels in 1924 and U.S. Patent No. 1845947 issued to Weisner and Hill in 1932 describe two such mechanisms which produce polygraphic substitution. The latter, that of Weisner and Hill, is of particular interest because it is based on a rather simple mathematical process which can yield true polygraphic encipherment for polygraphs of any size. The underlying mathematical process, invented by Prof. Lester S. Hill of Hunter College and described in the "American Mathematical Monthly" in 1929 (Vol. XXXVI, p. 306) and 1931 (Vol. XXXVIII, p. 135), is treated briefly, below; a more detailed treatment is contained in Appendix 8, "Lester S. Hill algebraic encipherment", which also includes remarks on the cryptanalysis of this method of encipherment.

(1) Since Professor Hill's system is mathematical in nature, the first step in its use involves the conversion of the plaintext letters into numbers by means of a conversion alphabet which shows a correspondence between the 26 letters of the alphabet and the 26 numbers from 0 to 25, such as the following:

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
0	9	3	5	24	6	18	8	11	1	21	14	15	12	4	10	25	17	7	19	20	2	22	16	23	13

(2) The numbers obtained through the conversion of the plaintext letters are next treated arithmetically through the application of algebraic linear functions, this treatment being performed by means of mod 26 arithmetic.[43] The numerical results yielded by the algebraic treatment are then converted back into letters by means of the conversion alphabet, to yield the cipher equivalent of the original plain text.

(3) For example, suppose that the message "NOTHING TO REPORT" is to be enciphered by trigraphs, and that, for this purpose, the enciphering keys[44] are 1, 2, 1; 5, 11, 3; 2, 4, 13. The message would be divided into trigraphs NOT-HIN-GTO-REP-ORT and the letters which result from the following operation would be taken as the cipher equivalent of the first trigraph:

Using the conversion alphabet in (1), above, (N O T) is converted into (12 4 19); then the foregoing keys are applied--

$$1\times12+2\times4+1\times19 = 12+8+19 = 13+1(26) = Z$$

$$5\times12+11\times4+3\times19 = 8+18+5 = 5+1(26) = D$$

$$2\times12+4\times4+13\times19 = 24+16+13 = 1+2(26) = J$$

Thus, $\overline{NOT}_p$ is enciphered as $\overline{ZDJ}_c$.

(4) A large number of sets of enciphering and deciphering keys can be constructed. It is even possible to construct keys which yield reciprocal encipherment, and it is this possibility which makes practicable the construction of a machine or device to accomplish the enciphering and deciphering.

[43] Using "mod 26 arithmetic", one considers as the sum or product of two numbers, the number from 0-25 which is obtained by subtracting 26 (or a multiple of 26) from the ordinary arithmetical sum or product of the numbers.

[44] Encipherment of polygraphs containing n letters requires the use of n^2 keys. Thus, 9 keys are necessary for trigraphic encipherment; digraphic encipherment requires only 4 keys, whereas tetragraphic and pentagraphic encipherment necessitate the use of 16 and 25 keys, respectively. The numbers selected for use as keys must be chosen according to rather definite rules based on the "theory of determinants"; otherwise, cryptographic ambiguity may result when decipherment is attempted. Appendix 8 contains more on this matter.

i. Attention is called here to the applications of Table 13 ("Four-square individual frequencies") of Appendix 2; this table has been reproduced here for convenience. If the cryptanalyst has at hand a fairly large volume of cipher digraphs produced by encipherment with a normal four-square, he may use Table 13 as an aid in placing the initial letters and final letters of the cipher digraphs into the appropriate cells of the cipher component sections on the basis of their uniliteral frequencies. Thus, if a distribution made of the initial letters of cipher pairs in a particular example shows Q_c, I_c, and C_c to be the letters of predominantly high frequency (listed in descending order of frequency), and if the distribution of the final letters shows F_c, Q_c, and P_c as the letters of predominantly high frequency (in descending order of frequency), these letters may be tentatively placed into a skeleton four-square matrix as follows (Fig. 70), based on the locations of the highest frequencies as given in Table 13:

(Table 13, Appendix 2)

[Based on a count of 6,000 digraphs]

		P_1					C_1		
A	B	C	D	E	244	225	375	394	197
F	G	H	I J	K	125	98	193	271	95
L	M	N	O	P	229	199	188	350	251
Q	R	S	T	U	148	162	258	427	295
V	W	X	Y	Z	42	12	34	91	97
212	317	358	308	249	A	B	C	D	E
120	108	216	256	85	F	G	H	I J	K
216	140	152	435	269	L	M	N	O	P
206	121	306	364	284	Q	R	S	T	U
38	29	21	147	43	V	W	X	Y	Z
		C_2					P_2		

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

Figure 70.

j. In attempting to diagnose the underlying cryptosystem in any particular polygraphic cipher, the student may gain some assistance from the following recapitulation:

(1) In digraphic ciphers the majority of repetitions will be an even number of letters apart and these repetitions should for the most part begin on the first letters of pairs and end on the last letterε of pairs. The majority of repetitions in trigraphic ciphers will be some multiple of three letters apart and these repetitions should for the most part begin on the first letters of trigraphs and end on the last letters of trigraphs.

(2) Digraphic ciphers may be revealed as such by the digraphic phi test, with additional support being given by the digraphic blank-expectation test; the presence of a null letter at the beginning of the cipher text might be disclosed by applying the two foregoing tests to a distribution of the digraphs which are formed when the first letter of the text is omitted.

(3) If either the uniliteral frequency distribution for the initial letters or for the final letters of the digraphs in a cryptogram exhibits monoalphabeticity, the cryptogram is probably a pseudo-digraphic cipher involving a large table of the type in Fig. 47 or 48. If both of the foregoing uniliteral frequency distributions reflect monoalphabeticity, the cryptogram may involve the use of a table of the type in Fig. 69.

(4) If the "decipherment" of a cryptogram by means of a four-square matrix containing four normal alphabets yields two monoalphabetic substitutions--one for the initial letters and one for the final letters of the pseudo-decipherment--the cryptogram may be assumed to be an inverse four-square cipher.

(5) If an ocular inspection or statistical evaluation of the cipher text of a cryptogram reveals a large number of "transparencies", the cryptogram probably involves a two-square system.

(6) If a cryptogram contains several cipher doublets, all of which are broken up when the cipher text is divided into digraphs, the cryptogram may well involve normal Playfair encipherment.

(7) If the cipher text of a cryptogram exhibits any invariable affinity of one of the letters J, K, Q, X, or Z for vowels (or, for that matter, another cluster of 5 or 6 letters), the cryptogram probably is in a small-matrix system employing sections consisting of more than 25 letters.

k. If a particular four-square cryptogram involves the use of a matrix in which either the plain component sections or the cipher component sections are normal alphabets, the matrix will be recovered through cryptanalysis in its original form, even when the components which are mixed have been derived by a transposition method or by no method at all. In Playfair cipher solution, the matrix can be recovered in its original form as long as the original matrix has been mixed in some systematic manner. However, in the case of two-square solution, there is no guarantee that the matrix can be recovered in its original form unless the original matrix has been keyword-mixed; if the original has been transposition-mixed, for example, the matrix which has been recovered through cryptanalysis--while being cryptographically equivalent to the original--will undoubtedly involve a permutation of the rows and columns of the original.

l. When four-square systems are encountered in which the matrix consists of four differently-mixed sections, reconstruction of the matrix is accomplished in a manner similar to that used in the analysis of two-square ciphers. If the sections are composed of keyword-mixed sequences, the original matrix may be recovered. Otherwise, the reconstructed matrix will in all probability be a permutation of both the rows and the columns of the original matrix, and there may be no way of recovering or or proving the original matrix.

m. In passing, it might be well to mention that any two-square system can be solved as a four-square system in which the matrix is composed of four mixed sections; upon the realization, from phenomena in the matrix reconstruction, that a two-square matrix is involved, the proper conversion can then easily be made.

(BLANK)