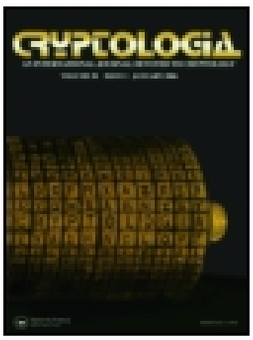


Cryptologia

ISSN: 0161-1194 (Print) 1558-1586 (Online) Journal homepage: http://www.tandfonline.com/loi/ucry20

PURPLE REVEALED: SIMULATION AND COMPUTER-AIDED CRYPTANALYSIS OF ANGOOKI TAIPU B

Wes Freeman , Geoff Sullivan & Frode Weierud

To cite this article: Wes Freeman , Geoff Sullivan & Frode Weierud (2003) PURPLE REVEALED: SIMULATION AND COMPUTER-AIDED CRYPTANALYSIS OF ANGOOKI TAIPU B, Cryptologia, 27:1, 1-43, DOI: 10.1080/0161-110391891739

To link to this article: http://dx.doi.org/10.1080/0161-110391891739

Published online: 04 Jun 2010.

PURPLE REVEALED: SIMULATION AND COMPUTER-AIDED CRYPTANALYSIS OF ANGOOKI TAIPU B

Wes Freeman[1], Geoff Sullivan[2], and Frode Weierud[3]

ADDRESS: (1) 2527 Mardell Way, Mt View CA 94043 USA. wesf@worldnet.att.net; (2) 64 Tennyson Road, Headless Cross, Redditch, Worcs., B97 5BJ UNITED KINGDOM. geoff@blueangel.demon.co.uk; (3) Le Pre Vert, 1041 Rte de Mategnin, F-01280 Prevessin-Moens FRANCE. Frode.Weierud@cern.ch.

ABSTRACT: PURPLE was the designation given by U. S. cryptanalysts to the cipher machine used by the Japanese Foreign Office for secure communications before and during WW 2. We present the structure and internal wiring of the machine, as well as details of the keying procedures and a system of abbreviations, which was used in the messages. We have also written a computer simulation of the PURPLE machine. Operation of the simulator is demonstrated by deciphering portions of the 14-part message delivered to the United States on Dec. 7, 1941. Finally, an automated cryptanalysis method for the PURPLE system is presented.

KEYWORDS: Pearl Harbor, PURPLE, RED, Rowlett, Stepping Switch, Computer Simulations, Hill Climbing, *Roma-ji*, *Romazi*.

INTRODUCTION

On December 6, 1941, the Japanese Government sent a message, which was divided for transmission into 14 parts (the 14-part message), to its embassy in Washington. An accompanying message directed that the 14-part message be delivered to the United States Government on December 7 at 1:00 p.m., which corresponded to 7:30 a. m. at Pearl Harbor, Hawaii. The 14-part message, which broke off negotiations between Japan and the United States, was enciphered on the PURPLE machine.

The 97-*shiki-obun In-ji-ki* (Cipher Machine 97) was also referred to as *Angooki Taipu B* by the Japanese. When cryptanalysis of the PURPLE machine began, the American cryptanalysts also referred to it as the "B Machine." However, it did not seem prudent to use the Japanese designation so a 'cover name' of PURPLE was adopted [5, p. 142]. The allied cryptanalytical services also used

trigram designators, so-called short titles, for the various cryptographic systems under study. PURPLE was given the short title JAA, with the sub-designations JAA-1 and JAA-2 for variants of the PURPLE keying method.

The operation of the PURPLE machine has been described in detail in [1-3]. Therefore, only a general overview of the machine will be presented, as well as specific details regarding the internal wiring of the machine.

The PURPLE machine was a substitution device, which replaced each plaintext letter with a ciphertext letter (or vice versa) from a permuted or 'scrambled' alphabet. The alphabet permutation changed after each letter was enciphered or deciphered, producing a polyalphabetic substitution system. Input and output to the PURPLE machine was provided by two conventional electric typewriters. Messages to be enciphered or deciphered on the PURPLE machine were entered on the keyboard of the input typewriter and the resulting ciphertext or plaintext was printed on the output typewriter.

The primary cryptographic element of the PURPLE machine is a 25 position stepping switch (also known as a stepping relay or uniselector). This device connects an input terminal to one of 25 output terminals. An electro-magnet, attached to the switch, advances the switch to its next position (e. g. from output 1 to output 2) in response to an electrical pulse. When the magnet is properly connected to the electric typewriter circuit, the stepping switch will advance to the next consecutive position each time a key on the typewriter is pressed. After reaching position 25, the next pulse will return the switch to position 1.

The stepping switch was originally developed in response to a need to automate the routing of telephone calls. An early U. S. patent (No. 447,918) for a stepping switch was granted to Almon B. Strowger on March 10, 1891. This device, with one input connected to the originating telephone, selected one of ten levels of switches, and then one of ten contacts on the selected level. This Strowger switch could, therefore, route a telephone call to one of 100 possible receivers. Over time, Strowger switches were developed with a wide variety in the number of levels and the number of contacts per level. One model, which is important to the discussion of the PURPLE machine, had six levels of switches and 25 contacts per level. Each level had a separate input, so that six inputs could simultaneously be connected to one of 25 outputs. Figure 1 shows a similar switch with eight levels and 25 contacts.

For the purposes of this discussion, it is important to understand that the internal components of the PURPLE machine which we present are actually the components that were used in the U. S. Army Signal Intelligence Service (SIS) recreation, or analog, of the original machine. The SIS cryptanalysts never actually saw the Japanese machine, and, as far as has been revealed in the

literature, only a few parts of one machine were ever recovered at the end of World War II.

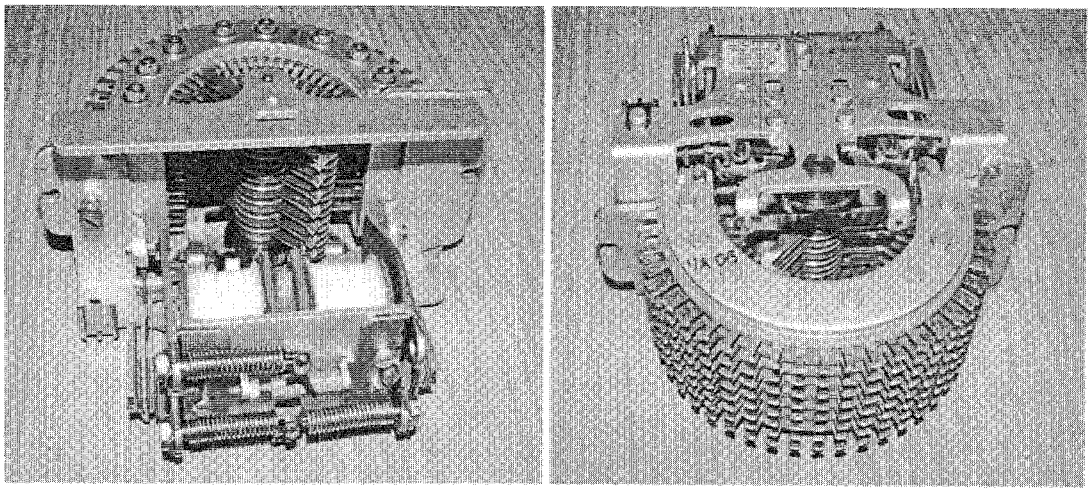


Figure 1. Two views of an eight level, 25 position stepping switch, showing the electromagnet that advances the switch. This particular example includes a second electromagnet which steps the switch in the reverse direction, but this feature was not required for the PURPLE machine.

BLOCK DIAGRAM

A simplified wiring diagram of the PURPLE cryptographic unit is shown in Figure 2[1]. The cryptographic elements are an input plugboard, stepping switches to permute the text, and an output plugboard. The design of the PURPLE machine is such that the input and output pluggings can be unrelated. However, Japanese practice was to have the plugging sequences identical or related[2]. In fact, the 'PURPLE ANALOG No. 1' on display in the U. S. National Cryptologic Museum at Ft. Meade, MD, has only one physical plugboard (only one plugboard is required if the plugboard connections are identical and double-contact jacks and plugs are used).

As can be seen from Figure 2, the 26 alphabetic characters (A-Z) are divided into one group of six and another group of twenty characters. These two groups are designated the sixes and the twenties, respectively, and represent the English

[1]Adapted from: NARA. RG457. NSA Historical Cryptographic Collection. *Schematic Drawings for the PURPLE Machine*. NR. 3204, Box 1017. The original drawing labels the twenties switches as "Special Sel" and the sixes switch "Master-6," and implies the plugboard shown in Figure 3.

[2]NARA. RG457. NSA Historical Cryptographic Collection. *Tentative List of Enigma and other Machine Usages*. NR. 1417, Box 580

vowels (AEIOUY) and consonants (BCDFGHJKLMNPQRSTVWXZ). However, the PURPLE machine did not specifically encipher vowels as vowels and consonants as consonants. (Further information on this arrangement will be presented later, in the "Regarding the 6-20 Alphabet Division" section.)

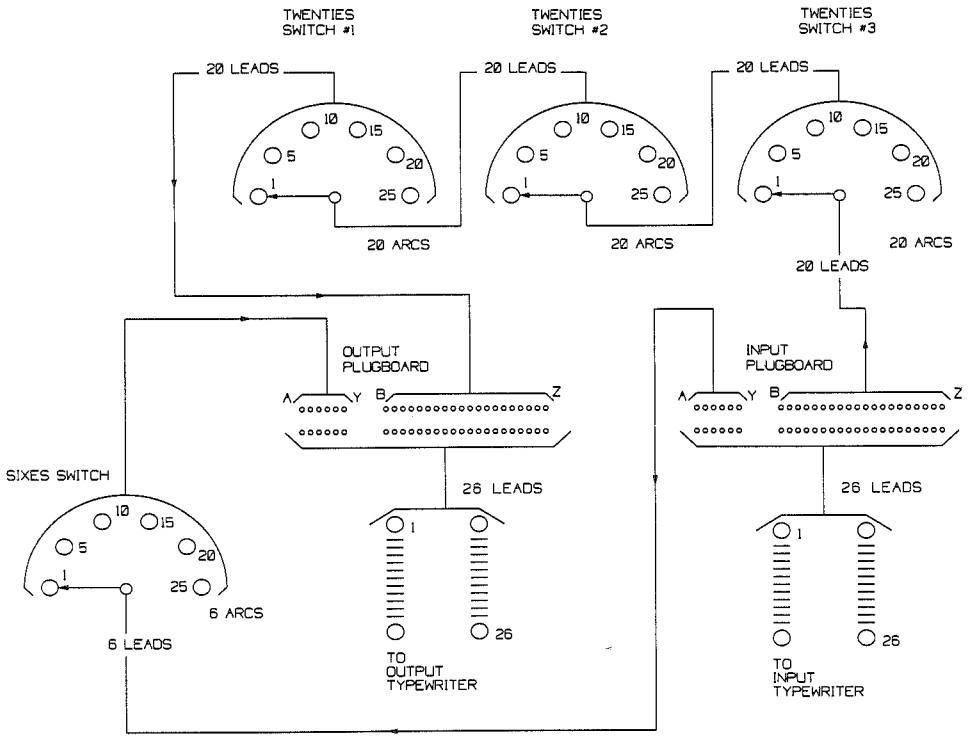


Figure 2. Simplified wiring diagram of the PURPLE deciphering unit.

The input jacks of the sixes portion of the plugboard are connected to the inputs of a six-level, 25 position stepping switch. The outputs of the switch are wired so as to permute (or rearrange) the input letters. As the switch advances through its 25 positions, the output is 25 scrambled alphabets (out of a possible 6!, or 720 alphabets). Unlike a rotor machine such as the ENIGMA, the 25 alphabets of the sixes are completely unrelated.

The input jacks of the twenties portion of the plugboard are connected to the inputs of a 20-level, 25 position stepping switch. As with the sixes, this switch is wired to produce 25 unrelated alphabets. However, the outputs of the first twenties switch are connected to a second switch and then in turn to a third switch. As will be shown shortly, the result is a twenties cycle of $25 \times 25 \times 25$, or 15,625.

Each twenties switch was actually constructed from four six-level switches, since 20-level, 25 position switches were not available. All four sections moved as one. The sixes switch included a second six-level section which also moved in parallel. The second section of the sixes switch controlled the stepping of the twenties switches and also the lamps which indicated the switch position.

The outputs of the stepping switches, either sixes or twenties #1, are connected to the output plugboard. This plugboard is then connected to an output typewriter, where the plaintext or ciphertext is printed.

It is important to note that during encipherment or decipherment any sixes letter was only replaced by itself or another sixes letter. Similarly, twenties letters were only substituted among themselves. This feature was very important in the cryptanalysis of PURPLE, as will be shown in the "Regarding the 6-20 Alphabet Division" section of this paper.

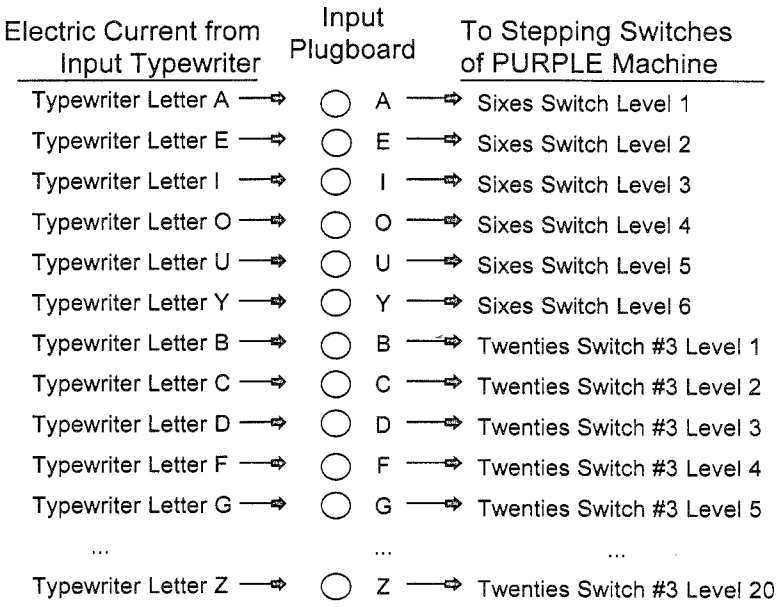


Figure 3. Wiring of the PURPLE machine's input plugboard. As an example, if the "B" wire from the typewriter is plugged into the "E" input of the machine, then the letter "B" will be enciphered by level 2 of the sixes switch. If "B" is plugged into the "F" input, on the other hand, then "B" will be enciphered on level 4 of the twenties switches.

The simplified diagram of Figure 2 shows a plugboard wherein two sets of jacks are located on the PURPLE analog machine itself. Individual patch cords or jumper wires would then be used to make the plugboard connections. Another possible arrangement is shown in Figure 3, where individual wires from the

typewriter are inserted into one set of jacks in the PURPLE machine. Either method will produce a correct result, and it is impossible to know which system the Japanese used.

Figure 3 also supplies more detail of the connections between the input typewriter, the plugboard, and the stepping switches. Any letter coming from the typewriter (i.e. on the left side of the diagram) can be connected to any input of the PURPLE machine (on the right). Therefore, pressing a key on the typewriter may result in that letter being enciphered either through the sixes switch or through the twenties switches. Selecting the alphabet permutation is covered in "The PURPLE Keying System" section of this paper.

Notice that the plugboard connections shown above are alphabetic, while the stepping switch connections (Level 1 etc.) are numeric. This nomenclature follows the SIS convention shown in the switch wiring tables (Figures 6-9). Assigning numbers instead of letters to the switches is a logical choice because the plugboard allows any letter to replace any other letter.

The physical layout of the Japanese plugboard is unknown, since no complete machine was ever recovered. Whatever the physical layout was, however, the wiring of the plugboards of the Japanese machine and the SIS analog were different [7, p. 2-1A]. The respective electrical (not physical) layouts of the plugboards were:

Japanese

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

SIS Analog

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

Because of the difference in plugboard layouts, the Japanese plugboard alphabet sequence (which is part of the daily PURPLE key) must be modified for use with the SIS analog machine. This operation is a simple substitution (for example, the letter in the 'V' position of the Japanese alphabet would be plugged into the 'B' socket of the SIS analog). In this paper, we will use the SIS alphabet.

Figure 2 shows the input typewriter connected to Switch #3 while the output typewriter is connected to Switch #1[3]. These are the correct connections for deciphering a normal PURPLE message. The input and output connections would be reversed when enciphering the message. However, messages could also be enciphered in 'reverse' mode (i. e. enciphered in decipher mode and vice versa).

[3]The U. S. Army Signal Intelligence Service (SIS) cryptographers numbered the switches as #3, #2, and #1, while the Japanese order was #1, #2, and #3. This would not have been obvious from cryptanalysis, but was only discovered because the Japanese occasionally sent new switch settings in messages which were themselves enciphered in the PURPLE system. Hereafter the SIS numbering will be used.

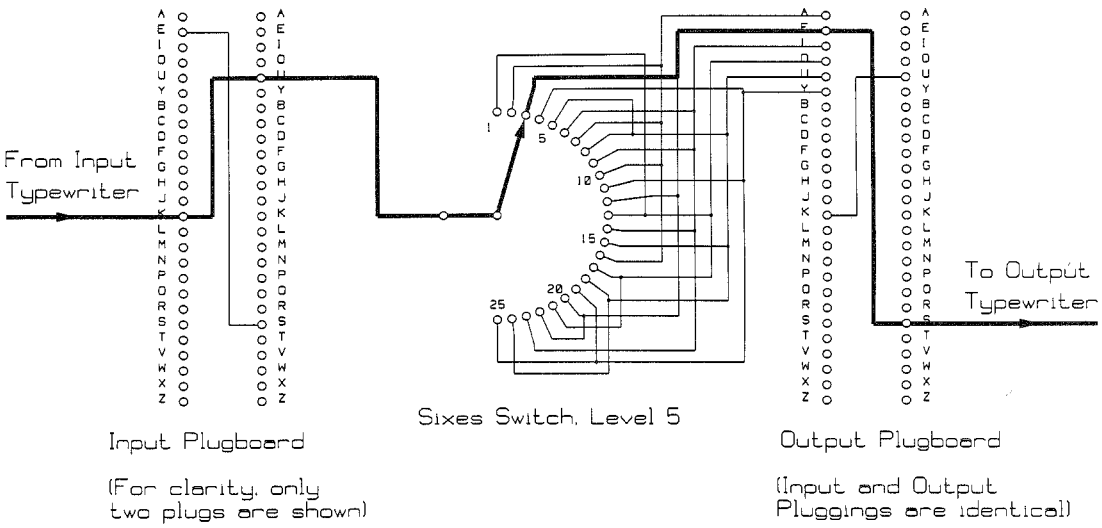


Figure 4. The electrical path through the PURPLE analog for a sixes letter.

Not shown in Figure 2 is the wiring of the switches, which required almost 2,000 wired connections. The sixes switch has six levels with 25 output contacts per level. The wiring table for the sixes switch is shown in Figure 6, where the top row denotes the input connection and each of the succeeding 25 rows denotes the permutation output at that switch position. For each of the 25 output levels, all of the "1" outputs of the switch must be connected together. The same rule applies to outputs 2 through 6. Then each group of outputs at each level must be wired together. With six levels and 25 contacts, this requires a total of 150 wires. Figure 4 shows the wiring of level 5 of the sixes switch and gives an example of the wiring complexity. The remaining levels of the sixes switch would show similar complexity. The twenties switches have 20 inputs and 25 positions, which requires 500 wires for each of the three switches. The result was what has been described as a "rat's nest" of wiring surrounding each switch [2, p. 232].

To understand the operation of the PURPLE machine, consider the circuit of Figure 4 with electric typewriters connected to the input and output. When a key on the input typewriter is pressed, for example "K," an electric current is applied to the "K" input of the plugboard. Depending on the plugboard substitution, "K" may be changed to one of the sixes, such as "U." Since "U" is the fifth vowel, the typewriter current will be applied to level 5 of the sixes switch. If the sixes switch is in position 3, then Figure 6 tells us that input 5 is connected to output 2 (which then connects to "E"). Next the output plugboard connects "E" to some other letter, perhaps "S," and finally the current causes an "S" to be printed on the output typewriter.

If the plugboard converts the input letter to one of the twenties letters, the operation of the machine is similar but more complicated. The input typewriter current must pass through Switch #3, then Switch #2 and finally Switch #1 before proceeding through the output plugboard and on to the output typewriter.

After each letter is enciphered or deciphered, the sixes switch and one of the twenties switches will move to the next position and create a new permutation. The result is a polyalphabetic system with a period of 25 for six letters and 15,625 for the remaining twenty letters.

The path through the PURPLE machine is therefore dependent on the switch permutations, the switch movement, and the alphabet plugboard substitution. A specific example of deciphering an actual message will be presented in the "Deciphering a PURPLE Message" section. First, however, the switch wiring and movement portions of the cryptographic system will be examined in more detail.

SWITCH WIRING TABLES

The wiring of the PURPLE switches is shown in Figures 6 through 9 [7, pp. A5-2, A4-4-6]. Each column represents one level of the stepping switch and each row is the permutation of the inputs at that position. For example, an electric current applied to the input of level 1 would appear at output 2 when the sixes switch is in position 1 and at output 6 at position 2. As previously mentioned, the switch inputs are connected to the alphabet plugboard, so input 1 of the sixes switch, Figure 6, is connected to plugboard letter A, input 2 to E, etc. For twenties switch #3, Figure 9, input 1 connects to B, etc.

One interesting feature of the switch wiring is that position 20 of Switch #1 is the identity permutation (1=1, 2=2, etc.). This cannot be a statistical anomaly since, as will be shown shortly, the switch wiring appears to be carefully chosen. One possibility is that this position was part of a test function of some sort, although no combination of switches #2 and #3 produce a corresponding identity permutation.

Another feature which should be noted is the balanced distribution of multiple-letter connections in the switches. For example the sixes switch has six inputs but 25 outputs, so each input must go to more than one output. Looking at column 1 of Figure 6, we find that input 1 is connected to output 1 at positions 3, 15, 19 and 22. Input 1 is also connected to outputs 2 through 5 at four positions, but is connected five times to output 6 (at positions 2, 7, 17, 21 and 23). Examining the remaining columns of Figure 6 reveals that each input level is connected to one output level five times, and that each output level has five connections once. A chart of the connections for the sixes, Figure 5, makes the

wiring symmetry more obvious. Note that each row and each column contains exactly one position with five output connections.

	Number of Connections to Each Output					
Input Level	Output 1	Output 2	Output 3	Output 4	Output 5	Output 6
1	4	4	4	4	4	5
2	5	4	4	4	4	4
3	4	4	4	5	4	4
4	4	5	4	4	4	4
5	4	4	4	4	5	4
6	4	4	5	4	4	4

Figure 5. The grouping of output connections for the sixes switch.

Switch	Sixes switch input (1=A, 2=E, etc.)					
Position	1	2	3	4	5	6
1	2	1	3	5	4	6
2	6	3	5	2	1	4
3	1	5	4	6	2	3
4	4	3	2	1	6	5
5	3	6	1	4	5	2
6	2	1	6	5	3	4
7	6	5	4	2	1	3
8	3	6	1	4	5	2
9	5	4	2	6	3	1
10	4	5	3	2	1	6
11	2	1	4	5	6	3
12	5	4	6	3	2	1
13	3	1	2	6	4	5
14	4	2	5	1	3	6
15	1	6	2	3	5	4
16	5	4	3	6	1	2
17	6	2	5	3	4	1
18	2	3	4	1	5	6
19	1	2	3	5	6	4
20	3	1	6	4	2	5
21	6	5	1	2	4	3
22	1	3	6	4	2	5
23	6	4	5	1	3	2
24	4	6	1	2	5	3
25	5	2	4	3	6	1

Figure 6. Sixes stepping switch - Decipher mode.

The twenties switches have 20 inputs but 25 outputs, and they show similar regularity. Looking at Figure 7, for example, we find on twenties switch #1 that input 1 is connected to output 17 at positions 3 and 9 and to output 16 at positions 7 and 23. Outputs 1, 5 and 12 are also "doubled." For all three of the twenties switches, every input is connected to every output at least once, and every input is connected to two outputs exactly five times. Furthermore, output 17 of switch #1 has doubled connections at inputs 1, 6, 9, 17 and 18. Again, every output of every twenties switch has exactly five doubled connections.

As previously mentioned, the PURPLE machine will operate in either the 'normal' or 'reverse' mode. It should be noted, however, that the switch connections shown in Figures 6 through 9 are for the decipher mode. For enciphering messages with the switch tables, the reciprocal of each position must be used. For example, at position 2 the connections for the sixes switch for enciphering would read 5, 4, 2, 6, 3, 1 instead of 6, 3, 5, 2, 1, 4.

Switch	Twenties switch #1 input (1=B, 2=C, etc.)																			
Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	6	19	14	1	10	4	2	7	13	9	8	16	3	18	15	11	5	12	20	17
2	4	5	16	17	14	1	20	15	3	8	18	11	12	13	10	19	2	6	9	7
3	17	1	13	6	15	11	19	12	16	18	10	3	7	14	8	20	4	9	2	5
4	3	14	20	4	6	16	8	19	2	12	17	9	5	1	11	10	7	13	15	18
5	19	6	8	20	13	5	18	4	10	3	16	15	14	12	7	2	17	11	1	9
6	2	11	9	14	7	19	6	3	18	13	12	8	10	15	16	17	20	4	5	1
7	16	7	6	18	9	10	13	1	17	2	5	4	11	19	20	14	8	15	3	12
8	1	20	7	16	12	14	5	18	15	10	13	6	8	3	4	9	11	17	19	2
9	17	9	11	8	20	18	7	14	1	16	15	5	19	2	6	12	4	10	13	3
10	12	8	17	9	3	20	4	10	14	5	7	18	2	16	13	6	1	19	15	11
11	20	1	16	11	2	17	9	4	8	15	10	13	3	18	14	5	6	7	12	19
12	5	4	15	2	13	19	6	16	12	14	8	7	17	10	18	3	9	1	11	20
13	15	17	10	19	16	2	11	8	9	7	3	14	18	13	12	1	5	20	6	4
14	11	12	7	3	8	15	16	6	4	20	2	5	1	9	19	18	10	14	17	13
15	12	16	2	7	4	8	15	19	5	1	11	9	20	17	6	14	13	3	18	10
16	8	15	18	1	12	11	17	14	20	16	13	19	9	7	3	4	2	5	10	6
17	7	3	5	18	17	13	19	20	14	11	9	10	2	6	1	15	12	16	4	8
18	10	13	4	14	18	3	2	17	11	19	20	1	6	12	9	7	15	8	5	16
19	13	7	9	12	20	16	14	10	19	6	1	2	11	4	5	3	18	17	8	15
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	9	20	12	5	10	17	1	13	7	15	4	3	16	8	18	11	19	2	14	6
22	18	15	2	13	1	7	10	5	19	17	6	20	9	11	12	8	3	4	16	14
23	16	18	19	10	11	20	5	9	1	4	12	13	7	6	17	2	14	15	3	8
24	5	8	1	15	19	9	12	2	6	3	14	17	4	20	16	13	18	10	7	11
25	14	10	4	8	9	12	3	11	17	20	19	6	15	5	2	18	16	7	1	13

Figure 7. Twenties stepping switch #1 - Decipher mode.

Switch Position	Twenties switch #2 input (1=B, 2=C, etc.)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	15	9	1	5	17	19	3	2	10	8	11	18	12	16	6	13	20	4	14	7
2	12	6	15	2	4	9	8	16	19	17	5	11	20	7	10	18	1	14	13	3
3	4	18	5	8	16	1	12	15	20	14	13	17	11	2	7	9	6	3	10	19
4	6	11	2	20	14	7	18	12	15	3	8	5	10	1	17	19	9	16	4	13
5	7	2	13	3	9	4	17	14	1	12	18	20	6	11	16	15	5	8	19	10
6	5	17	14	7	10	9	19	20	8	13	1	2	16	3	4	12	11	18	6	15
7	8	4	3	11	19	13	2	9	12	16	10	17	14	15	20	6	18	1	7	5
8	20	1	16	10	15	8	14	11	18	5	3	7	13	17	19	4	2	9	12	6
9	9	8	7	15	5	2	4	13	17	1	11	6	19	18	14	10	3	20	16	12
10	10	12	11	18	8	16	20	17	5	6	9	3	4	19	13	7	1	14	15	2
11	11	7	14	4	18	20	6	1	13	19	12	15	5	9	16	2	17	10	8	3
12	2	3	9	10	13	14	15	16	7	11	20	12	18	6	1	5	8	17	19	4
13	16	10	15	1	17	3	13	9	4	7	6	8	2	14	5	11	12	19	18	20
14	19	16	18	12	3	13	9	10	6	2	17	14	11	4	7	20	15	5	1	8
15	18	14	12	19	1	7	10	6	11	15	5	9	8	20	17	4	3	13	2	16
16	20	3	19	2	4	5	11	14	9	10	18	16	15	12	8	7	13	6	17	1
17	3	6	4	14	2	12	16	5	18	20	7	19	1	15	9	8	10	11	13	17
18	5	15	20	9	10	17	1	19	13	12	4	2	7	6	11	14	16	8	3	18
19	14	20	13	17	5	18	8	4	2	15	16	1	9	19	3	6	7	10	12	11
20	8	11	1	6	19	14	5	18	17	3	10	13	12	20	15	16	4	2	7	9
21	17	19	6	1	12	15	20	7	16	9	3	11	13	10	2	18	8	4	5	14
22	1	5	12	20	6	11	14	8	9	7	19	4	3	13	10	17	18	16	15	2
23	16	8	10	13	11	6	19	5	3	4	15	20	17	2	18	1	14	7	9	12
24	19	13	8	16	20	10	7	1	2	18	14	6	9	5	12	3	17	15	11	4
25	13	1	17	15	7	4	16	3	14	5	2	10	18	8	11	9	19	12	20	6

Figure 8. Twenties stepping switch #2 - Decipher mode.

REGARDING THE 6-20 ALPHABET DIVISION

Dividing the alphabet into vowels and consonants was carried over from PURPLE's predecessor, *Angooki Taipu A* or the RED machine. When first introduced, the RED machine enciphered vowels (A, E, I, O, U, and Y) as vowels and consonants as consonants [5, p. 115]. The ciphertext thus produced was 'pronounceable' under the telegraphic rules in force at the time, and this was important because random text telegrams were charged twice as much as plain language or 'pronounceable' (but nonsense) text. For example, Boris Hagelin patented (U. S. Patent 1,484,477) a machine (CRYPTOCODE-TYPER, Model A 4), which would convert numeric code groups (for example "7342589016") into 'pronounceable' letter groups (such as "olavuxutip") [4, pp. 341-344]. The Japanese would have certainly been aware of this device, since they had purchased other cryptographic devices from Hagelin.

Switch Position	Twenties switch #3 input (1=B, 2=C, etc.)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	7	19	11	3	20	1	10	6	16	12	17	13	8	9	4	18	5	14	15	2
2	15	17	14	2	12	13	8	3	1	19	9	4	10	7	11	20	16	6	18	5
3	2	11	20	12	1	19	4	10	9	14	6	15	13	3	7	16	18	8	5	17
4	16	3	12	9	4	20	6	19	18	2	5	8	14	11	10	1	15	17	13	7
5	12	18	16	4	9	3	15	13	6	20	8	2	7	10	5	19	14	1	17	11
6	13	9	5	6	8	7	12	17	14	18	20	10	2	19	11	15	4	3	1	16
7	4	7	2	15	17	10	19	5	8	16	1	12	3	13	6	14	20	9	11	18
8	9	6	4	10	18	16	8	14	5	12	17	1	20	15	13	19	2	11	7	3
9	5	14	18	17	13	15	11	12	7	8	3	6	1	2	20	4	9	10	16	19
10	11	16	9	18	3	12	5	15	10	1	14	17	2	4	19	6	8	7	13	20
11	19	8	3	15	14	5	1	11	2	10	12	16	18	20	17	7	13	4	9	6
12	1	12	17	13	9	7	14	2	15	4	5	11	6	16	3	8	18	19	20	10
13	3	4	10	12	1	18	2	8	14	13	19	7	16	6	15	9	17	20	5	11
14	9	11	6	5	10	4	17	19	13	15	7	2	12	18	14	20	1	16	8	3
15	8	13	14	16	19	12	20	7	10	3	15	9	4	17	1	11	5	2	6	18
16	18	16	15	4	2	17	13	12	6	11	20	19	14	5	9	1	8	7	3	10
17	14	1	7	20	6	13	16	18	12	9	4	17	5	11	2	3	10	15	19	8
18	17	19	1	11	7	2	18	4	3	8	10	5	15	12	16	9	6	13	20	14
19	10	15	2	14	11	6	7	1	16	20	13	3	9	8	18	17	19	5	12	4
20	20	9	8	6	12	11	2	5	4	7	16	14	17	3	15	10	13	19	18	1
21	11	20	13	8	16	10	18	14	19	6	15	4	1	17	7	5	3	9	2	12
22	16	5	10	19	4	18	15	17	1	3	2	20	11	6	8	13	7	12	14	9
23	6	10	19	16	5	9	1	20	17	4	11	18	7	14	13	2	12	8	3	15
24	8	7	5	1	15	14	9	16	11	17	18	6	19	20	3	12	4	2	10	13
25	13	2	17	7	14	8	3	9	20	5	16	10	6	1	12	15	11	18	4	19

Figure 9. Twenties stepping switch #3 - Decipher mode.

However, the RED and PURPLE machines were entirely different in their cryptographic operation. RED utilized the Damm 'half rotor' with 26 contacts for slip rings on the shaft. These contacts were split into two groups of 20 and 6 contacts. These contacts were connected to a rotor of 60 positions (the least common multiple of 20 and 6) at the end of the shaft. A pinwheel controlled the movement of the rotor, advancing one, two, or three positions for each letter enciphered and with a variable period of 41 to 43 positions [2, pp. 218, 219].

Since the operating principles of the RED and PURPLE machines are entirely different, it is difficult to speculate on why the 6-20 division was maintained. The simplest explanation is that the designer or designers did not recognize this division as a critical weakness. Especially when vowels are only enciphered as vowels, as happened when RED was first introduced, the effect is startling. The distribution of vowels in Romanized Japanese language includes many bigrams, such as OO, UU, AI and IA, and even trigrams such as YUU and YOO [5, p. 117]. These combinations made it possible for cryptanalysts to guess at plaintext words and thereby recreate the operating principles of the machine.

The Japanese abandoned the vowel/consonant division a few months after the introduction of the RED machine, and thereafter any of the 26 letters could be plugged into the sixes [5, p. 123]. It seems logical to assume that Japanese cryptographers realized that mapping vowels to vowels reduced the security of the machine. Unfortunately, the damage was already done because SIS cryptographers had the earlier messages and used them to establish the operating principles of the machine. The new keying procedure simply meant that SIS had to develop methods to separate the sixes and the twenties. This was accomplished by analyzing the letter frequencies and noting that the sixes usually had a higher or lower frequency of occurrence than the twenties. With the decryption process established and the alphabets recovered, RED was deciphered continuously by SIS until the machine was removed from service in 1941. RED messages were also deciphered by the British, Germans, and Russians. [3, p. 82]

Although the operating principles of RED and PURPLE were entirely different, the 6-20 division gave SIS cryptographers a ready-made wedge into cryptanalysis of the PURPLE machine. Using the techniques already developed for RED, the sixes of PURPLE messages were identified and the encryption algorithm was revealed. Japanese messages in the RED system typically began with a message number. The U. S. Army's SIS kept careful records of these numbers, which gave them many ciphertext-plaintext pairs. The same numbering system was continued with the PURPLE machine, and this provided enough data to enable the recovery of the system for enciphering the sixes [5, p. 146].

With the sixes recovered, the underlying plaintext of some messages could be identified. Eventually enough plain text was recovered to reveal the switch motion and wiring of the twenties and the PURPLE machine was solved.

SWITCH MOTION

The motion of the sixes switch is straightforward, advancing one position after each letter is enciphered. When the sixes switch reaches position 25, the next letter enciphered will return the sixes switch to position 1. However, the position of the sixes switch does control the movement of the twenties switches.

For the twenties, one switch moves after each letter is enciphered. A 'fast' switch advances one position after each letter is enciphered, unless the sixes switch has reached the 25 position. In this case, the 'middle' switch advances while the 'fast' switch does not. If the 'middle' switch has also reached the 25 position, then the 'slow' switch will advance when the sixes reaches 24, followed (on the next encipherment) by the 'middle' switch. An example of the switch motion when the slow switch moves is shown in Figure 10.

Switch Positions			
Sixes	Twenties #1	Twenties #2	Twenties #3
21	1	25	5
22	2	25	5
23	3	25	5
24	4	25	5
25	4	25	6
1	4	1	6
2	5	1	6

Figure 10. Turnover of the twenties switches is controlled by the sixes switch. Note that only one twenties switch moves with each letter enciphered. For this example, the motion is: #1 is fast, #2 is middle, and #3 is slow.

Since only one of the twenties switches moves with each letter enciphered, $25 \times 25 \times 25$ letters can be enciphered before the cycle will repeat. Any of the three twenties switches can be the fast, middle or slow position, giving six possible switch movements.

THE STEPPING SWITCH AS A CRYPTOGRAPHIC ELEMENT

PURPLE and its naval cousins, CORAL and JADE, were unique in that they were the first, and perhaps the only, known cryptographs to use stepping switches to permute the plaintext. One feature of the stepping switch as a cryptographic element is that the alphabets which the switch produces are totally unrelated. This was a new concept for the SIS cryptographers. Frank Rowlett, Solomon Kullback and others had, under the direction of William F. Friedman, studied a variety of cryptographic machines, such as the Hebern, KRYHA, and Hagelin [5, ch. 6]. Rowlett and Kullback had even solved the RED machine in a ciphertext-only attack. All of these machines, however, had a cyclic operation of some sort. PURPLE's sixes were solved fairly quickly, since there were only 25 permutations, but no exploitable pattern could be identified for the twenties.

The output of the PURPLE machine is cyclical, of course, but Japanese keying methods, combined with the complexity of the machine, conspired to hide the patterns. Some plaintext could be identified through the deciphered sixes, such as English language messages that had been delivered to the United States State Department. Although the Japanese initially used only 120 different switch starting positions (out of 15,625 switch positions with six possible motions per position), they did use a different alphabet every day. This ensured that very few messages would be sent with exactly the same key (additional information on keying methods is presented in The PURPLE Keying System section of this

paper). Without matching cipher and plain text, in depth, the motion and wiring connections of the twenties switches could not be established.

Motion Indicator	Switch Positions #1	#2	#3	Motion	Input to Twenties B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Z
1	1	24	6	123	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	2	24	6		**D**	G	J	K	F	Z	B	C	R	M	V	H	P	W	Q	N	Y	S	L	T
	1	25	6		P	W	C	B	R	N	L	Z	J	T	F	M	H	V	Y	K	S	G	D	**Q**
	2	25	6		H	Q	Z	V	T	Y	K	L	S	N	B	R	F	J	G	W	M	C	P	**D**
2	1	24	6	132	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	2	24	6		**D**	G	J	K	F	Z	B	C	R	M	V	H	P	W	Q	N	Y	S	L	T
	1	24	7		N	C	D	T	G	P	K	V	H	R	Z	F	J	**Q**	L	M	B	Y	W	S
	2	24	7		Y	Z	P	N	C	H	W	J	F	T	L	B	S	**D**	K	R	V	G	Q	M
3	1	24	6	213	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	1	25	6		P	W	C	B	R	N	L	Z	J	T	F	M	H	V	Y	K	S	G	D	**Q**
	2	24	6		**D**	G	J	K	F	Z	B	C	R	M	V	H	P	W	Q	N	Y	S	L	T
	2	25	6		H	Q	Z	V	T	Y	K	L	S	N	B	R	F	J	G	W	M	C	P	**D**
4	1	24	6	231	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	1	25	6		**P**	W	C	B	R	N	L	Z	J	T	F	M	H	V	Y	K	S	G	D	Q
	1	24	7		N	C	D	T	G	P	K	V	H	R	Z	F	J	**Q**	L	M	B	Y	W	S
	1	25	7		S	N	H	K	Z	M	V	C	R	Q	D	L	G	**P**	B	J	F	W	Y	T
5	1	24	6	312	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	1	24	7		N	C	D	T	G	P	K	V	H	R	Z	F	J	**Q**	L	M	B	Y	W	S
	2	24	6		**D**	G	J	K	F	Z	B	C	R	M	V	H	P	W	Q	N	Y	S	L	T
	2	24	7		Y	Z	P	N	C	H	W	J	F	T	L	B	S	**D**	K	R	V	G	Q	M
6	1	24	6	321	**Q**	Y	V	L	H	C	F	G	M	S	B	P	D	K	W	T	N	J	Z	R
	1	24	7		N	C	D	T	G	P	K	V	H	R	Z	F	J	**Q**	L	M	B	Y	W	S
	1	25	6		**P**	W	C	B	R	N	L	Z	J	T	F	M	H	V	Y	K	S	G	D	Q
	1	25	7		S	N	H	K	Z	M	V	C	R	Q	D	L	G	**P**	B	J	F	W	Y	T

Figure 11. Permutations of the twenties letters for all switch motions.

The PURPLE machine has six switch motions, which affect the cryptographic process in two different ways [2, p. 239]. All six possible motions of the twenties switches are shown in Figure 11. The designations of the Motion Indicators, 1 through 6 in the left hand column, are those used by SIS to set the motion on

their PURPLE analog machines. The switch motion, e.g. 132, is given in fast, middle and slow notation (i.e. Switch #1 is fast, Switch #3 is middle and Switch #2 is slow).

The first two lines of each motion indicator are the consecutive decipherments of each of the twenties letters (i.e. only the fast switch has advanced one position). The third and fourth lines of each indicator are the identical letter decipherments except that they also include one step of the middle switch. The deciphered letters only represent the permutation of the twenties switches, since the plugboard was plugged 'straight through' (i.e. B=B, C=C, etc.).

Motion indicators 1, 2, and 4 show that the columns are permuted as the middle switch turns (in motion 1, for example, the letters "Q" and "D" under input B are moved to column Z when Switch #2 advances). Indicators 3, 5, and 6 show isomorphic substitutions (for motion 5, "D" exactly replaces "Q" when switch #1 advances). Even though the effects of the motion are either permutation or isomorphism, however, each of the six possible switch motions produces its own unique output.

One feature of PURPLE which is ignored in Figure 11 is the effect of the sixes switch. As mentioned previously, the middle switch will advance when the sixes switch advances from 25 to 1, while the fast switch does not move. Figure 12 shows a decipherment matrix for the twenties letters before, during, and after the sixes switch passes position 25. The motion, which produces these permuted columns, is 231 (Motion Indicator 4).

Two pairs of permuted columns are highlighted as examples in Figure 12. Notice that an apparently unrelated alphabet is inserted at position 4 (either the column PXKJ has become PXKC or JZXW has become TZXW). Another unrelated alphabet is at position 30.

What actually happens is that the columns of letters are changing in addition to being permuted. For example, the complete column of decipherments for the uppermost PXK text is:

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

When the middle switch moves, the column is transposed (to Input column R) but the fast switch does not move. Therefore the last letter of the previous column, J, becomes the first letter of the next column:

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

Note that the fifth, sixth and seventh letters deciphered under Input column R in fact are J, S, and H. Each time the middle switch advances, the letters in the column will descend one position.

When the slow switch moves, the order of the letters within the column is

permuted (scrambled). The effect of middle and slow switch movement on other switch motions is similarly complex. A simulator for the PURPLE machine will be presented in the "A Graphical PURPLE Simulator" section of this paper. This simulator may be used to explore the effects of switch motion more fully.

Switches				# of Letters	Input to Twenties																					
Sixes	#1	#2	#3	Enciphered	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Z
22	1	12	6	1	**P**	C	D	W	N	S	T	J	F	G	B	K	R	Z	V	H	L	Q	X	M
23	1	13	6	2	**X**	B	G	R	Q	D	J	T	W	Z	V	C	L	P	F	M	H	S	N	K
24	1	14	6	3	**K**	F	R	D	L	Q	W	S	B	M	J	X	N	H	G	C	T	P	Z	V
25	1	15	6	4	J	K	H	C	F	L	Q	R	V	D	N	S	W	X	M	G	Z	T	P	B
1	1	15	7	5	**Z**	L	W	G	R	S	X	H	F	B	P	Q	T	J	C	V	N	K	M	D
2	1	16	7	6	**X**	K	R	J	D	L	G	B	W	C	V	N	Z	S	M	T	H	Q	P	F
3	1	17	7	7	**W**	N	F	Q	L	V	D	X	M	J	R	Z	B	H	T	S	G	P	C	K
22	1	11	7	26	B	F	C	N	G	Z	J	P	H	X	K	S	W	M	V	Q	R	D	T	L
23	1	12	7	27	L	S	R	H	J	K	Z	D	N	M	X	T	Q	**P**	W	F	B	C	V	G
24	1	13	7	28	H	D	L	M	T	C	P	G	Q	K	N	J	S	**X**	R	W	V	B	F	Z
25	1	14	7	29	T	Q	N	C	S	X	H	R	L	V	Z	W	P	**K**	D	B	J	F	G	M
1	1	14	8	30	F	D	T	X	M	V	L	B	R	W	S	Z	J	C	K	H	N	G	Q	P
2	1	15	8	31	K	C	**Z**	S	D	B	F	V	H	Q	R	P	N	G	J	X	W	M	L	T
3	1	16	8	32	Q	M	**X**	L	F	C	W	T	B	N	D	V	H	J	S	G	R	P	K	Z
4	1	17	8	33	P	T	**W**	V	K	J	M	S	X	Z	L	R	G	Q	H	D	F	C	N	B

Figure 12. Cipher alphabets of the twenties as the middle switch advances. Note one unrelated alphabet after every 25 characters.

The appearance of the unrelated alphabet in Figure 12 is evidence of an interesting feature of PURPLE. An examination of all of the switch wiring tables reveals that no input is connected to the same output at two successive positions, even at the 'roll over' from position 25 to 1. Since only one switch moves at a time, the PURPLE machine will never encipher the same plaintext letter to the same ciphertext letter twice in succession. Unlike some other machines such as the Enigma, however, PURPLE will encipher a plaintext letter to the same ciphertext letter.

THE PURPLE KEYING SYSTEM

PURPLE messages utilize a 2-part key. The first part of the key provides the switch starting positions and movement, while the second part of the key defines the daily alphabet.

The switch settings were determined by a 5-digit indicator group at the beginning of each message. Initially, the indicator was selected from a list of 120[4].

[4]See Appendix A. This is not the original Japanese document. The indicators were recovered by

Sixty of the indicators had all odd numbers (e.g. 73159), while the remaining 60 had all even numbers (e.g. 60284). By 1 May 1942, an additional 120 indicator groups, including both odd and even numbers, had been added [7, pp. A2-A-4, A2-A-5].

To resist cryptanalysis, it is desirable that messages not be produced 'in depth.' Since PURPLE is a hard-wired machine, the number of starting positions must be limited in order to prevent overlaps when sending long messages. On the other hand, limiting the starting positions increases the chances that two or more messages will be sent with the same settings. To disguise the limited number of starting positions, the message indicator was enciphered by adding (with carry) a predetermined five-digit number. This additive was changed monthly. For example, PURPLE messages sent in November, 1941 used an additive of 82313; December, 1941 messages used an additive of 03210.

The indicator of the 14-part message was further disguised by first transposing the true indicator and then applying the additive group. For example, part 1 of the 14-part message was enciphered with the switch settings corresponding to indicator 15739. This indicator was enciphered as follows:

Message indicator	15739
Digit sequence	ABCDE
Transposition sequence	BEDAC
Transposed indicator	59317
December additive	03210
Encrypted indicator	62527

It should be noted that not all PURPLE messages used the 240 starting positions mentioned previously. Certain highly sensitive messages, sent between Tokyo and Berlin, carried the external designation HIKAL. The HIKAL messages derived their switch starting positions from the original list of 120, but modified the normal starting positions (for example by reading the switch positions down a column instead of across the row). HIKAL messages were also read by SIS, either by cryptanalysis or because the keying instructions were transmitted in PURPLE messages that were broken.

U. S. codebreakers often recorded the switch positions and movement using the format a-b,c,d-ef, where a is the starting position of the sixes switch, b, c, and d are the twenties switches #1, #2, and #3 respectively, e is the fast switch

cryptanalysis.

and f is the middle switch. For example, the indicator for part 1 of the 14-part message would have been written: 9-1,24,6-23. Another format for recording the indicator was a-b-c-d-n, where n represents one of the six possible switch motion indicators given in Figure 11. For this second case, the part 1 indicator would be 9-1-24-6-4, where 4 represents switch motion 231.

The PURPLE daily alphabets were derived from a list of 1,000 basic alphabets, which was arranged as 50 pages with 20 lines per page[5]. On any given day, all messages were enciphered using the same basic alphabet. However, the alphabet was permuted before being used[6]. There were several methods of selecting the alphabet, and these methods changed over time, but the method employed during December 1941 was as follows:

1. Add the month and day to get the page number.
2. Starting with the bottom line, count up one line for each day of the month. For days of the month beyond the 20th day, subtract 20 from the date. The resulting line contains the basic alphabet.
3. Apply the permutation to the basic alphabet. The permutation key was:

```
19 15 7 23 11 2 8 18 3 14 24 25 17 13 22 4 5 26 6 1 10 21 12 9 20
```

For December 6, 1941, the alphabet was selected as follows:

1. Month (12) plus day (6) selects page 18.
2. Day (6) selects the sixth alphabet from the bottom, which is line 15. The alphabet is:

```
SUQDIAKXZVYGMLOJRENFWPTHBC
```

3. Applying the permutation from the preceding paragraph gives the final alphabet. The permutation operation is a simple three-step procedure:
   a) Number the letters of the basic alphabet.
   b) Write the permutation key under the alphabet
   c) Transpose the letters. For example, the first transposition key is 19, and the 19^{th} letter of the basic alphabet is N, so the first letter of the final alphabet is N. (The transposition sequence is arranged to place the alphabet in the proper AEIOUYBCDFG...Z sequence). Therefore:

[5]See Appendix B. This page is not the original Japanese document. The alphabets shown were recovered by cryptanalysis.

[6]Also, some stations used a different permutation. In addition, the permutations were changed from time to time. The result was that the effective number of alphabets was much greater than 1,000.

```
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
 S  U  Q  D  I  A  K  X  Z  V  Y  G  M  L  O  J  R  E  N  F  W  P  T  H  B  C
19 15 07 23 11 02 08 18 03 14 24 25 17 13 22 04 05 26 16 06 01 10 21 12 09 20
 N  O  K  T  Y  U  X  E  Q  L  H  B  R  M  P  D  I  C  J  A  S  V  W  G  Z  F

Alphabet=    Sixes          Twenties
             AEIOUY     BCDFGHJKLMNPQRSTVWXZ
             NOKTYU     XEQLHBRMPDICJASVWGZF
```

When the U. S. Army's SIS began working on the cryptanalysis of PURPLE in 1939, the combination of 120 starting points and different alphabets every day severely limited the number of messages with identical keying elements. In addition, of course, the SIS could only decipher a small percentage of intercepted messages based on recovery of the sixes. The cryptanalysis of PURPLE was only successful when several deciphered messages, with the same indicator (and hence identical starting positions) but from different days (and therefore different alphabets), were exploited. SIS cryptanalysts actually converted all of the ciphertexts to the same alphabet, which then gave them enough messages in depth to map the switch connections of the twenties. [3, p. 92].

An interesting feature of the PURPLE machine may have aided SIS cryptanalysts in matching cipher and plain text. As previously mentioned, PURPLE will encipher a plaintext letter to the same ciphertext letter. However, the frequency with which this occurs is different for the sixes and the twenties. Examining the sixes deciphering table (Figure 6) at position 1 reveals that input 3 is connected to output 3 (i.e. the identity permutation). Since the input and output plugboards are identical, any letter enciphered at this position will be the identical plaintext letter. Further examination of Figure 6 shows that 25 out of the possible 150 locations have the identity permutation. Thus, 16.6% of the sixes will have identical encipherments.

The twenties situation is more complicated, because there are three switches in series. Logic would say that there is a 1 in 20 (5%) chance that two letters will coincide, and this is not far from the truth. Using a PURPLE simulator, the same letter was enciphered at each position in the twenties alphabet, for all 15,625 switch positions. The result was that the identity permutation exists at 15,610 out of 312,500 ($20 \times 25 \times 25 \times 25$) locations, or 4.99%.

Even if the suspected English language version of a PURPLE message were located, the two texts would not have matched exactly. The PURPLE machine did not encipher punctuation, so the Japanese used a variety of three-letter codes for punctuation, formatting instructions, etc. These codes (which are described in the next section of this paper) would not have appeared in the final plaintext message. With the sixes already deciphered, however, and with 16.6% of the

sixes identity permutations, adjusting the English text to match the ciphertext would have been greatly simplified.

FREQUENCY ANALYSIS OF SOME PURPLE MESSAGES

One of the first steps in cryptanalyzing a PURPLE message is to separate the alphabet into the sixes and twenties groups. Being a polyalphabetic system, one would expect that the frequency distributions would be fairly smooth but that the sixes distribution would vary from the twenties depending on how many high or low frequency letters were in each group. In practice, however, this method is not as simple as it seems. An extreme example was found in a message of 565 letters, sent from Washington to Tokyo on November 10, 1941[7]. The ciphertext letter distributions for this message are shown in Figure 13.

The letters in the sixes alphabet are marked with an asterisk, and their distributions range from 14/565 to 34/565. When deciphered, this message was found to be an enciphered code message whose 'plain text' was actually a Japanese code designated 'CA.' The CA code used two letters to represent words, phrases, numbers, punctuation, etc. The structure of some code groups effectively doubled the contribution of that group to the frequency count (for example, parentheses are always used in pairs and AX="(" while XA=")"; also, numerals are represented by doubled letters). Code groups with high usage, such as HE, which appeared 15 times in the message, further skew the distribution. Figure 14 illustrates some CA code groups and their associated meanings[8].

Q	X	H	T	R	O	S	A	E	M	V	D	I
37	34*	29	27*	26	25	25	24	24*	24	24	23	22
U	W	B	G	K	P	Z	J	C	F	L	Y	N
22*	21	19*	19	19	19	17	16	15	15	14*	13	12

* Sixes letters

Figure 13. Ciphertext frequency distribution for the message of November 10, 1941.

Code Group	AX	XA	ZA	AZ	YY	EE	HE
Meaning	(	)	A*	Z*	1	2	*no* (of)

*Letters (A-Z) were used to spell words not in the codebook

Figure 14. Some representative Japanese CA code groups and their meanings.

[7]NARA. RG457. NSA Historical Cryptographic Collection. *Diplomatic Messages From/To Tokyo*. Nr. 1802, Box 733.

[8]NARA. RG457. NSA Historical Cryptographic Collection. *Cryptanalytic Working Aids for Japanese Cipher*. NR. 2047, Box 772

While the CA code message is perhaps an extreme example, Japanese language messages sent in the PURPLE system did not necessarily have 'normal' plaintext distributions. The Japanese Foreign Office employed a three-letter code for common words and phrases. This code was actually a type of shorthand, in that many of the 'code groups' were simply Japanese words with the vowels removed[9]. The effect of this code was to alter the plaintext frequency distribution as well as reduce the message size. In addition to words and phrases, numbers were represented by two or three letters. Examples of these 'shorthand' code groups are shown in Figure 15[10].

The message text was further condensed because the bigrams *ai* and *ei* were replaced by the single letters L and X respectively. In addition, the Japanese Foreign Office used the *roma-ji* form of writing instead of the more common Hepburn system[11]. The *roma-ji* system, for example, shortened the Hepburn *kana shi* to *si*, *chi* to *ti*, and *tsu* to *tu*. (In addition, *ji* was replaced by *zi* and the hyphen was not transmitted, so the word *roma-ji* would itself be written as 'romazi.') The effects of these two contractions can be demonstrated on the English word 'ambassador,' which in Japanese is *taishi*. Replacing *ai* with L and *shi* with SI produces the final plaintext group TLSI.

Code Group	BTN	DQN	STM	TKB	BB	DD	MS	ZJR
roma-ji Word	*betuden*	*daizin*	*setumi*	*tokubetu*				
Hepburn Spelling	*betsuden*	*daijin*	*setsumei*	*tokubetsu*				
Meaning*	separate wire (telegram)	minister	explanation	special	1	2	3	20

* The exact meaning of many Japanese words is context-dependent

Figure 15. Examples of code groups which reduce the vowel count in PURPLE messages.

In addition to words and numbers, PURPLE messages utilized a three-letter code for punctuation. Typical groups were LFL for comma and CFC for period. Words in a foreign language were denoted by FAE (begin foreign spell) and FIO

[9]In its structure, the Japanese code was similar to the Phillips Code. Developed by Walter P. Phillips, the Phillips code was an unclassified commercial code which was developed to speed the transmission of news reports sent by telegraph (Mr. Phillips worked for Associated Press and United Press). Representative Phillips Code groups would include ABD=aboard and CKT=circuit.

[10]NARA. RG457. NSA Historical Cryptographic Collection. *Purple Machine Abbreviations and Garble Chart*. NR. 2839, Box 952.

[11]NARA. RG457. NSA Historical Cryptographic Collection. *Introductory Lessons In The Study Of Telegraphic Japanese*. NR. 1480, Box 589.

(end foreign spell). Letters C, F, L and X are common in the punctuation codes (in fact, F is in every group). The addition of punctuation codes further skews the expected letter frequency distribution. A matrix of punctuation codes is shown in Appendix C.

In spite of the contractions and punctuation, it is easier to identify the sixes in part 1 of the 14-part message (see Appendix D) than in the 'CA' code message, since the content is mostly English text. Even so, the distribution, shown in Figure 16, is not definitive for all the letters. Note that two letters, H and I, are more numerous than the actual sixes letter T, while ten letters exceed the sixes letter Y.

O	K	U	N	I	H	T	E	F	W	D	R	X
82*	70*	64*	61*	57	55	54*	53	50	50	49	47	47
A	M	L	Y	C	J	Z	P	V	B	G	S	Q
46	44	43	43*	42	41	40	39	39	38	38	34	31

* Sixes letters

Figure 16. Letter frequency distribution for part 1 of the 14 part message.

However, each station employing the PURPLE system used the same daily alphabet (derived using the rules given previously in the PURPLE Keying System section) for every message sent that day. Therefore, several messages, if they are from the same originator and sent on the same day, can be combined in a frequency analysis. When ciphertext for the first seven parts of the 14-part message is combined, the identity of the sixes is more readily apparent (Figure 17).

O	T	K	U	N	Y	E	D	H	I	G	W	L
422*	384*	382*	379*	367*	338*	315	306	304	298	295	293	287
V	F	Z	B	M	J	A	C	X	S	R	P	Q
285	284	284	282	275	269	268	267	266	257	249	237	237

* Sixes letters

Figure 17. Letter frequency distribution for parts 1 through 7 of the 14-part message.

DECIPHERING A PURPLE MESSAGE

To demonstrate the functioning of the PURPLE machine, decipherment of the first three letters of a message will be demonstrated. The message selected is part 1 of the 14-part message (see Appendix D). The keying elements for this message are:

```
Sixes=9
Switch #1 = 1
Switch #2 = 24
Switch #3 = 6
Motion=231 (fast, middle, and slow, respectively)
Alphabet  = Sixes          Twenties
            NOKTYU   XEQLHBRMPDICJASVWGZF
Plugboard = AEIOUY   BCDFGHJKLMNPQRSTVWXZ
Stepping                      11111111112
Switch    = 123456   12345678901234567890
```

The first three ciphertext letters of the message are ZTX.

The first letter is a twenty, so the input plugboard will connect "Z" to "X" (and then to the twenties stepping switch input 19). Referring to the decode table for Switch #3, we find that at position 6 input 19 is connected to output 1. From the decode table for Switch #2, at position 24 input 1 is connected to output 19. The decode table for Switch #1 shows that at position 1 input 19 is connected to output 20. Finally, the output plugboard connects output 20 ("Z") to the letter "F," which is the plaintext letter. After this letter is deciphered, the fast switch (Switch #2) and the sixes switch will advance to their next respective positions.

The second ciphertext letter, "T," is one of the sixes. Decipherment is similar to a twenties letter, except that there is only one sixes switch. The sixes input plugboard converts "T" to "O" (and then to input 4 of the sixes switch). The sixes switch has advanced to position 10, so input 4 is connected to output 2. The output plugboard connects output 2 ("E") to "O," which is the plaintext letter. Succeeding letters are deciphered in a similar manner by incrementing the switch positions in accordance with the motion indicator. (Plaintext FOV is the punctuation code for 'open parenthesis'. See Appendix C)

The machine switch positions and paths for the first three letters of the message are shown in Figure 18.

Cipher Letter	Input Plugboard	Twenties Sw #3	Twenties Sw #2	Twenties Sw #1	Sixes Switch	Output Plugboard
Z	Z>X>19	Pos 06 19>1	Pos 24 1>19	Pos 01 19>20	Pos 09	20>Z>F
T	T>O>4	Pos 06	Pos 25	Pos 01	Pos 10 4>2	2>E>O
X	X>B>1	Pos 06 1>13	Pos 01 13>12	Pos 01 12>16	Pos 11	16>T>V

Figure 18. The path through the PURPLE machine when deciphering the first three letters of the message shown in Appendix D.

THE 14-PART MESSAGE

To demonstrate the operation of the simulator, as well as highlight the task facing U. S. cryptographers, the first 216 letters of part 1 of the 14-part message (see Appendix D) have been deciphered. The decipherment and translation are presented on four lines, as follows: cipher text (CT), plain text (PT), garble correction and expanding the *roma-ji* spelling to Hepburn (HS), and final translated text (FT). The deciphered text is show in Figure 19. Explanations of the text are contained in tihe translation notes.

As mentioned in the translation notes (see Figure 19), part 1 appears to begin with several nonsense words. To explore this possibility further, parts 2 and 3 of the 14-part message were also deciphered. Both of these parts show unintelligible letters at the beginning of the message. It is interesting to note, however, that the first eight letters of part 2 are all located on the middle row of the English-language keyboard, while the first six letters of part 3 are found on the top row of letters.

Beginning plaintext of part 2 of the 14-part message[12]:

```
PT: HJKLFDSA FYX KZ ZR DD FOV JYN NO DD FOV FAE HOWEVER
HS:          FYC*
FT: HJKLFDSA  #   9  0  2  (   14 of  2  (  BFS However
```

Beginning plaintext of part 3 of the 14-part message:

```
PT: UREOPWNVBSKM XFC KZ ZR DD FOV JYN NO MS FYX FAE NEVER
FT: UREOPWNVBSKM  #   9  0  2  (   14 of  3  )  BFS Never
```

Random text at the beginning of messages was apparently not used when PURPLE was first introduced. It appears not to have been a common practice. A more frequent subterfuge was bisecting a message, which is explained in the next section.

Three cipher groups appear to be missing from the NARA copy of part 1 of the 14-part message (Appendix D). The group count is 265 (GR265) while there are only 262 groups in the file copy and the final 10 cipher groups are themselves incomplete. In addition, there is no indication in the deciphered text that this is a multipart message. However, the last two groups available yield plaintext of ITALYC - - - OV, which probably represents ITALY CFC FOV. The missing "one of fourteen" instruction, JYN NO BB FYX, was probably located in the last three groups.

[12] FYC may be a radio intercept garble or a coding error. Note the two 'open parenthesis' later in this line.

```
CT: ZTX ODNWKCCMAV   NZ      XYWEE        TU QTCI     MN
PT: FOV TATAKIDASI   NI      MUIMI        NO MOXI     WO
HS:  (  TATAKIDASHI  NI      MUIMI        NO MOJI     WO(3)
FT:  (  Type out(1)  into(2) meaningless  of letter(s)

CT: VEU VIWB    LUA  XRR TL  VA  RG  NTP  CNO         IUP
PT: IRU BESI    FYX  XFC KZ  ZR  DX  OOV  BTN         FYX
HS: IRU BESHI  )     XFC KZ  ZR  DD  FOV  BTN         FYX
FT: be  shall  )(4)  #   9   0   2   (    betuden(5)  )

CT: JLC    IVRTPJKAUH VMU   DTH KTXYZE  LQTVWG BUH FAW  SHU
PT: FAE    MEMORANDUM FIO   FOV OOMOJI  BAKARI FYX RAI  CCY
HS: BFS(6) MEMORANDUM EFS(7) (  OOMOJI  BAKARI  )  RAI  CCF
FT: BFS    MEMORANDUM EFS    (  capital only    )  next paragraph
                             letters

CT: LBF BH EXM YHF LOWD-KWHKK NX EBVPY HHG HEKXIOHQ
PT: LFC BB CFC THE GOVE-NMENT OF JAPAN LFL PROMPTED
HS:               GOVERNMENT
FT: sub  1  .  The Government of Japan  ,  prompted
    para-
    graph

CT: HU H WIKYJYH PPFEAL NN AKIB OO ZN FRLQCFLJ TTSSDDOIOCVT-
PT: BY A GENUINE DESIRE TO COME TO AN AMICABLE UNDERSTANDIN-
HS:                                           UNDERSTANDING
FT: by a genuine desire to come to an amicable understanding

CT: ZCKQ TSH XTIJCNWXOK UF NQR -TAOIH WTATWV
PT: WITH THE GOVERNMENT OF THE -NITED STATES
HS:                           UNITED
FT: with the Government of the United States
```

Figure 19. Decipherment and translation of the opening portion of part 1 of the 14-part message.

Translation Notes for Figure 19:

1. *Tatakidashi*. In this context, it means "type [bat] out (a letter)".
2. *Ni*. In, at, to, here: into.
3. *Wo*. Japanese fill word. No English translation.
4. The sentence in the parentheses does not really make sense. Even rephrased to "Shall be typed out into letters of no meaning" it does not fit with the rest of the message. As it comes before the message number (# 902) it is possible that it could be random text meant to hide stereotype beginnings, hence "Text with no meaning".
5. BTN. *Betuden (roma-ji)* — *betsuden*, which means "another telegram" or "additional telegram".
6. FAE. Code group for "Begin Foreign Spell" — BFS.
7. FIO. Code group for "End Foreign Spell" — EFS.

THE EVOLUTION OF PURPLE KEYING METHODS

On August 20, 1941, all users of the PURPLE system were instructed to divide messages into two, three, four or five parts and transmit the parts in irregular order [7, p. 2-37]. For example, a sender might bisect a message by beginning encipherment in the middle of the plain text. At the end of the original message the sender would insert "DDDD" to indicate the end of part two of two parts. Encipherment would then continue from the beginning of the original message and proceed to the bisection point. Finally, "DDBB" was inserted to indicate the end of part one of two parts and random letters added (if necessary) to complete the last five letter group.

This change eliminated the cribs derived from message numbers. However, by this time (actually by April 1, 1941) all of the 120 original starting points and about 50% of the alphabets had been recovered [7, p. 2-1A]. Thus, many messages could be read simply by using a known key. If cryptanalysis was required to recover the alphabet, the "DDDD" and "DDBB" groups provided useful cribs. The message contained in Appendix E is an example of a bisected message.

By July 1943, the Japanese had developed four different rules for the security of PURPLE messages[13]. Each of these rules was assigned a separate trigram

[13]NARA. RG457. NSA Historical Cryptographic Collection. *Systems Survey (1943)*. NR. 2022, Box 772.

designator by Allied cryptanalysts:

1. JAA, the regular traffic.
2. JAB, super secret messages of state.
3. JAC, code (or cipher) instructions.
4. JAD, the HIKAL traffic from Berlin to Tokyo.

All four of these rules used the same PURPLE hardware, however, and all were broken. Since new keying instructions were sent in the current system, the new keys were compromised as soon as they were transmitted. A PURPLE analog machine had also been sent to Britain, and there was a daily exchange of keys with the Japanese Diplomatic Section of GC & CS, Berkeley Street, London.

With the keying system identified and an analog machine completed, U. S. cryptanalysts were able to decipher PURPLE messages almost as rapidly as the Japanese recipients were. A study[14] of PURPLE message processing delays in November 1944, requested by Frank Rowlett, examined 100 messages selected at random. The deciphering time averaged 1.5 hours, but eliminating four messages (which had garbled indicators or text and resulted in delays of 17, 12, 12, and four hours) reduced the average decipherment time to 42 minutes.

Over time, however, changes were introduced to the keying system that created at least temporary problems. The HIKAL messages between Tokyo and Berlin have already been mentioned. Messages on a circuit between Tokyo and Moscow, labelled JAA-1 by U. S. cryptanalysts, were first intercepted on September 10, 1944[15]. The keying indicator for JAA-1 was five letters instead of five numerals. The starting points for the switches were derived from the position of the letters of the keying indicator within the message alphabet. The daily alphabet was selected using the method previously described, except that the alphabet below the basic alphabet was used as the permutation key (that is, if line 9 in Appendix B was the basic key then line 10 would be the permutation key). After a few messages were broken, the encoding algorithms for the alphabet and starting positions was discovered and further messages were read with little delay.

A much more serious problem arose in April 1945, with the introduction of another new keying system[16]. The new system, which was named JAA-2 by

[14]NARA. RG457. NSA Historical Cryptographic Collection. *Japanese Cryptanalysis Material, 1942-1945*. NR. 2777, Box 948.

[15]NARA. RG457. NSA Historical Cryptographic Collection. *Rules for PURPLE Machine and JAA-1 Instructions*. NR. 3127, Box 1004.

[16]NARA. RG457. NSA Historical Cryptographic Collection. *Signal Security Agency General Cryptanalytic Branch - Annual Report FY 1945*. NR. 4360, Box 1380.

Allied cryptanalysts, used the indicator groups from an obsolete Japanese code (identified as JBA) to define new starting positions for each PURPLE message. At that time, finding the starting positions for a message using hand methods took about one week. This delay might have been acceptable for compiling the initial 120 indicators, perhaps, but it was far too long if new starting points had to be found for every message. Clearly the process had to be speeded up, so Rapid Analytical Machine (RAM)[17] equipment was employed. In one case, an IBM card reproducer was connected to the PURPLE analog. Generating and printing all of the PURPLE starting positions on IBM cards with this machine took about one day, but the resulting card file reduced the delay in finding the switch starting positions from one week to 15 minutes.

AUTOMATED CRYPTANALYSIS OF THE PURPLE MACHINE

One cipher text-plain text pair from the PURPLE machine was found without a given key (see Appendix E). This was seen as an opportunity to determine the key using modern computer techniques, using the cipher text only. The intercepted message consisted of five parts; four being of 50 five-letter groups and one of 28 five-letter groups. The message was enciphered in one run with a single starting key. It should be noted that the division of the message into five parts only applies to the cipher text, and was probably done for efficiency in transmission. The plain text was bisected before encryption, however, in accordance with the previously mentioned instructions.

Breaking this cipher consists of finding:

i. The sixes and twenties alphabets.

ii. The stepping motion order of the twenties switch bank.

iii. The start positions for the twenties and sixes switch banks.

The sixes alphabet can often be found by simple frequency analysis. However for this message, the plain text of which is Japanese *kana* with some 3-letter code groups, this is not easily done. The letter distribution is given in Figure 20.

E	K	U	H	G	N	J	M	T	S	C	V	B
57	56	56	54	51	51	48	48	48	47	46	45	42
I	Y	D	R	O	Z	W	A	L	F	P	Q	X
42	42	41	41	40	39	38	36	36	31	31	25	24

Figure 20 Letter frequency distribution for the message of 27 July 1942.

[17] A general description of some RAM devices, although not mentioning PURPLE applications, is included in [1, pp. 359-362].

There is no obvious group of six letters as suitable candidates for the sixes alphabet. It could be EKUHGN or ALFPQX but the counts are not outstanding. We could, of course, simply ignore the tactic of finding the 6-20 split, and do a "brute force" attack on the entire machine. However, the theoretical key space that we must deal with is daunting: 25 starting positions for the sixes and 25 x 25 x 25 for the twenties, six different twenties motions, and 26! alphabets (assuming we do not know the alphabet permutation). These variables total $25 \times 25 \times 25 \times 25 \times 6 \times 4 \times 10^{26}$, or 9.45×10^{32}. Clearly a more sophisticated approach was required, and the method selected was hill climbing.

AN INTRODUCTION TO HILL CLIMBING

Finding the alphabet of a substitution cipher is not necessarily an "all-or-nothing" affair. Deciphering a message with a key which is 'almost' correct can allow portions of the plaintext to be recovered. In the English language, for example, such common bigrams as 'er' and 'th' and trigrams like 'ion' and 'ing' will become visible. If we can somehow measure how close to 'true' plaintext a trial decipherment is, we can then modify the trial key and move toward the 'true' key. The method we will use to modify the trial key is called hill climbing[18].

Assume we have a hypothetical cipher system with a four-letter key. The key can take any value from AAAA to ZZZZ. Different keys will produce differing amounts of plaintext, but there is only one correct key. In addition, we have ciphertext, enciphered with key TRVJ, but we do not know the key. Finally, we can decipher the message with any key and measure the percentage of plaintext. The graph of the relationship between trial alphabet and plaintext for our unknown message is shown in Figure 21.

Rather than starting with key AAAA and trying all possible alphabets in a brute-force approach, we begin with a key roughly 1/4 of the way between AAAA and ZZZZ, such as FFFF. Deciphering the message, we get a result, which we will call a score, of roughly 40% plaintext. Next, we try the two adjacent keys, FFFE and FFFG. Scoring these two keys shows that FFFE gives a higher score, so we discard FFFF and make FFFE the new trial key. Then we try the two keys adjacent to FFFE, and continue this procedure until the score does not increase.

At this point we have climbed the hill and reached a peak score, but it is obvious that there is a problem with this simple example: our test has reached a peak at DQVJ, which is not the correct key. We must, therefore, try several different starting points, following each one to a peak, before we can expect to find

[18]A detailed description of Jim Gillogly's hill climbing can be found in the Cipher Challenge archive `http://groups.yahoo.com/groups/CipherChallenge`.

the correct answer. To prevent any relationship between the keys, which might result in our climbing several related but incorrect hills, the different starting points should be selected randomly.

In practice, it may be more practical to consider the graph of trial keys versus plaintext as 3-dimensional rather than the two dimensions of Figure 21. Consider the key space as a topographical map, where altitude represents the percentage of plain text[19]. We can now modify the key in several directions, pick the result that yields the highest slope, and thus climb the hill more quickly.

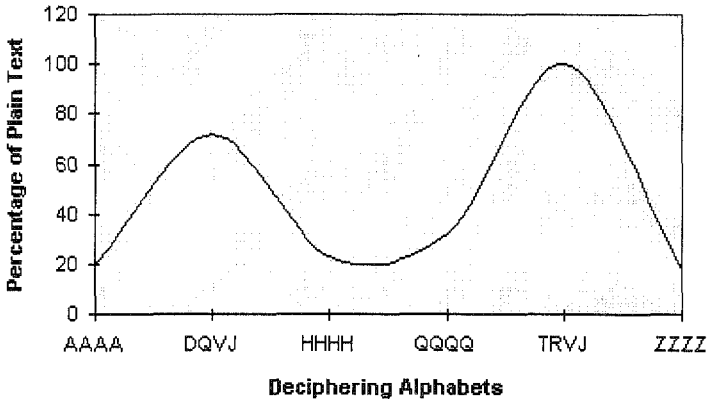


Figure 21. The relationship between message key and plain text for a hypothetical cipher system.

IDENTIFYING PLAIN TEXT IN A PURPLE MESSAGE

The simplest approach to cryptanalysis of the PURPLE message is to blindly start a hillclimb of the entire machine and test for possible Japanese text. However, if we can divide the problem into two parts, and solve the sixes and twenties separately, we can significantly reduce the keyspace that we have to test. Identifying plaintext when only the sixes are properly deciphered may be difficult, but this difficulty is offset by the fact that there are only six letters and 25 starting positions. Knowing the sixes starting position will help to decipher the twenties by identifying the proper stepping position for the middle and slow switches. Knowing the stepping position is important because the twenties, with three switches and six motions, will require far more hillclimbing that the sixes. Therefore it is more convenient to determine the sixes alphabet and sixes switch setting first.

[19]Gillogly, Jim. 1995. Shotgun Hill-Climbing. *The Cryptogram*. LXI(6): 12-13.

The hill climbing attack on the message needs a system to measure the amount of correct plain text in the decrypted message. The simplest system to apply is the Index of Coincidence (IC) [6, p. 68]. The SIS documentation gives an expected IC of 0.0554 for abbreviated Japanese. The actual IC score for our plain text message is 0.0578 and for the corresponding cipher text 0.0392. A hill climb applied only to the sixes alphabet using IC scoring may not give good discrimination, since the decrypt will only contain a small amount of correct text at this stage. Another score system was sought with higher discrimination. The plain texts of five other PURPLE messages from 1941 were available to construct bigram and trigram frequency tables. These five messages were also *kana* and included some 3-letter code words. Some examples of the most common plain text bigrams and trigrams in the five messages are given in Figure 22. Some examples of 3 letter code groups and their meaning are given in Appendix C. Using the frequency tables obtained from the five messages, the score for a trial decrypt is determined by summing the log of either the bigram or trigram frequencies found in the output text [6, p. 77].

KUN	NAR	KAN	ASI	SUR	URU	SIN	ARI	IWA	UNO
17	16	15	15	15	15	14	14	13	13
SEN	IKA	ISI	YOR	KUS	DEN	RYO	NNI	ATA	AWA
13	12	12	12	12	11	11	11	10	10

KU	EN	SI	RU	NO	WA	AN	AR	NI	KO
62	57	56	49	48	48	47	46	44	44
NA	NO	UN	IS	IT	IK	KI	ON	YO	AI
41	40	37	36	36	35	34	32	31	30

Figure 22. Frequency of 20 most common bigrams and trigrams in the sample plain texts of total length 2220 characters.

SIXES ALPHABET AND SWITCH POSITIONS

Since the sixes alphabet is only enciphered by the single sixes switch, this alphabet and the switch position can usually be found at the same time. The hill climb starts with a random alphabet of length 26. The first six letters being assigned to the sixes alphabet and the rest to the twenties. The sixes start position is first set to position 1. The twenties settings are ignored for this stage and can be set to any motion and any start position. The message is decrypted in the usual way using the six and twenty split. The letters passing through the twenties switches will be random since it is unlikely that the twenties settings are correct at this time. The letters passing through the sixes switch will eventually start to give correctly decrypted characters in the hill climb when the correct sixes switch setting is reached. A score is determined by summing the log of

bigram frequencies found in the output text. Since the decrypt will only contain fragments of the plain text, due to the 6 and 20 split, the bigram score is the most appropriate system to use for this stage.

After obtaining a start score, the next adjacent alphabet key is selected. This is achieved by swapping each letter pair in turn, in the full 26 alphabet. A trial decrypt with the new alphabet is then made. The swapped pair is retained if the score increased, otherwise the swapped pair is restored. For example, if the first random alphabet is:

```
FWKAQJHNLUBPTSZMYRXDOCVEIG
```

we then swap the first two letters:

```
WFKAQJHNLUBPTSZMYRXDOCVEIG
```

and decrypt and score the message. Assuming this gives an increase in score, we retain the swap and then try the next pair swap:

```
KFWAQJHNLUBPTSZMYRXDOCVEIG
```

This continues until all pair swaps have been done, restoring any that did not increase the score. After all letter pair swaps are tested, for any increase in score the key setting is written to a log file. A new random alphabet is then generated, the swapping and scoring process is thereby repeated 100 times. This entire process of 100 hill climbs is then repeated for each of the 25 possible sixes switch start positions. The log file of bigram frequency score, sixes switch position and sixes alphabet records the highest scores. For the unknown key message we obtain the sixes log file:

Score	Six	Alphabet	Score	Six	Alphabet
2484	00	VYIZPK	2794	02	DQXLIM
2531	01	GOYZBU	2826	04	QZPLXI
2586	01	GOYFQU	2831	04	YXQLFI
2595	01	EOYLJU	2836	11	AFBVPW
2621	01	OESLJU	2838	11	AFBVPW
2623	01	OESLJU	2905	13	XPILRQ
2635	01	EOHJLU	2931	13	XPILRQ
2677	01	CNSBMK	2976	13	XPILRQ
2728	01	QMPZLF	2989	13	XPILRQ
2763	02	LZPAWB	3006	13	XPILRQ
2780	02	FQELIX			

The highest score of 3006 occurred for switch start position 13, with a sixes

alphabet XPILRQ. These settings were thus taken as the sixes settings and used in the next stage to find the twenties settings.

It will be noted that several sixes alphabets occur more than once in the log file. This is due to the 100 random restarts allowed for the hill climb for each sixes switch setting. An increase in the score during the run of 100 can occur due to a different arrangement of the twenties letters, which are random at this point.

The number of trial decrypts used to determine the sixes settings is:

$$\frac{26!}{24! \times 2!} \times 100 \times 25 = 812,500 .$$

We can compare this with the theoretical key space for the sixes settings, assuming we are free to select any six letters for the sixes alphabet:

$$\frac{25 \times 26!}{20!} \approx 4.14 \times 10^9 .$$

TWENTIES HILL CLIMB

The score system used for the twenties hill climb uses the trigram frequency table since this gives more discrimination and gives a faster hill climb. Bigrams were used for the sixes hill climb because the twenties letters of each trial decryption were random. Recognizing trigrams with only six out of 26 letters known would have a very poor chance of success. For the twenties hill climb, however, the sixes letters will all decipher correctly so scoring trigrams is appropriate.

The twenties hill climb attack requires $25 \times 25 \times 25 \times 6$ separate hill climb operations to be performed to find the twenties alphabet, since every switch setting must be examined. Obviously this could take some time unless the hill climb program is optimised and a suitable rule applied for deciding when to give up with a hill climb and then move on to the next key setting. For this purpose, one of the other *kana* messages of similar length with a known key was examined. Ten random start hill climb operations were run with the machine set to the correct switch positions for the message. Each hill climb was terminated if the score did not increase for two consecutive random re-starts. The score was recorded at this point. This procedure was repeated for consecutive non-increasing re-starts of length 4, 8, 12, 16 and 20. With 12 consecutive non-increases as the end point, nine of the ten hill climbs achieved a score close to the maximum possible. At 16 this had increased to all ten. It was therefore decided to abandon a hill climb if the score did not increase for 16 consecutive re-starts. This ensured that the run time would not be excessive and give a good

chance of finding the key for the unknown message. Using this rule, the random re-starts rarely exceeded 50 and most of the incorrect settings were abandoned quickly.

TWENTIES HILL CLIMB RESULTS

With the empirically determined hill climb rule, the program was run separately for each switch motion. Four personal computers were available for this task, so four switch motion orders could be set running: 312, 321, 213 and 231. After several hours[20], the PC running motion 321 produced the successful log file:

Score	Switch position	Alphabet
314	321 13 - 01, 06, 11,	XPILRQ ZGOKUCSMJNFTYAEVWHDB
327	321 13 - 01, 21, 19,	XPILRQ FDTUAHSCZGYBWKVNMJOE
337	321 13 - 22, 01, 22,	XPILRQ ABDUYWOSFZHVKGTNMCJE
1959	321 13 - 25, 23, 12,	XPILRQ BYECZVNOUDFKGTWSHAMJ
2421	321 13 - 25, 23, 12,	XPILRQ FYECZVNOUDBKGTWSHAMJ

A trigram score of 2421 is a good indication that the switch settings have been successfully found and that the alphabets are correct, or very nearly correct. The correct key is switch motion 321, sixes switch set to 13 and twenties switches set to 25, 23, 12. The twenties alphabet being FYECZVNOUDBKGTWSHAMJ.

With 20 letters in the pair swap process, the typical number of message decrypts to hill climb the twenties is:

$$\frac{50 \times 20!}{18! \times 2!} \times 25 \times 25 \times 25 \times 6 \approx 8.9 \times 10^8 .$$

This is a significant improvement over a brute-force approach. Given a sixes alphabet and starting position, the possible key space[21] for the twenties is:

$$20! \times 25 \times 25 \times 25 \times 6 \approx 2.28 \times 10^{23} .$$

A GRAPHICAL PURPLE SIMULATOR

A graphical computer simulation of the analog PURPLE cipher machine is available for download from the Author's Crypto Simulation Group WWW home

[20] The hill climb software was written using the C programming language. It was prepared with a view to fast development rather than attempting to run fast. It could probably be optimised to give a significant reduction in run time.

[21] Calculated with a Keuffel & Esser 4053-3 slide rule.

pages:

`http://www.blueangel.demon.co.uk/crypto`

`http://home.cern.ch/frode/crypto/`

This computer simulation program is designed to operate on Intel based PCs under Microsoft Windows operating systems. Readers may like to use this to decrypt the PURPLE message given in Appendix D.

In addition to encipher and decipher functions, the simulator also includes features to simplify cryptanalysis of PURPLE messages. The identity alphabet (A=A, B=B, etc.) can be loaded with a few clicks of the mouse. Any or all of the twenties switches can be 'frozen' so that the permutations of the machine (e. g. Figures 11 and 12) can easily be studied, and the switch settings can be changed rapidly with the mouse.

Another debugging tool is the sequence table generator. This feature provides a table of the complete substitution alphabets (sixes and twenties) for each of the first 50 switch motions of the machine. When the switch starting positions and alphabet are known, the sequence table can be used to decipher a PURPLE message by hand. If the starting positions are known but the alphabet is not, the sequence table can be used as an aid in recovering the alphabet [7, p. 2-9]. This process requires known plaintext, and consists of finding letter relationships in the sequence table that match the ciphertext/plaintext pairings. As previously mentioned, early PURPLE messages began with a message number (e.g. FYC DD BB), and often referenced another message, so guessing the first 20 or more characters was not unreasonable. Readers can also attempt to place the bisection point within the message of Appendix E by creating a sequence table beginning at switch positions 13 - 1, 6, 3 - 321, which are the switch positions at group 41 of the ciphertext (group 41 is "HYHAJ"). The bisection occurs somewhere within groups 41-50 and the plain text is "DDDD XFC GWHQHQ" (2 of 2, Msg #588). The message also contains plaintext "DDBB" in the last two groups.

CONCLUSION

The PURPLE machine, with its use of stepping switches as the primary cryptographic element, was a unique device. However, it retained the 6-20 alphabet split of its predecessor, the RED machine. This feature allowed U. S. Army cryptanalysts to identify enough matching ciphertext and plaintext to eventually recover the wiring of the machine. Once the wiring and keying methods were recovered, U. S. personnel could decipher PURPLE messages, such as the 14-part message, as quickly as their Japanese counterparts.

Although reading the 14-part PURPLE message could not prevent the United

States' entry into World War II, or indeed the Pearl Harbor attack, a great deal of information was gleaned from the PURPLE intercepts. The influence of intercepted PURPLE messages was even felt in the European theater, where significant information on German defences in Normandy was obtained by reading the messages of General Oshima Hiroshi, the Japanese Ambassador to Germany. General Oshima toured the Normandy region in 1943 and his detailed report, enciphered with the PURPLE machine but deciphered by Allied cryptanalysts, aided planning for the D-Day operations [1, p. 318].

Appendices A through E of this paper will permit the reader to decipher actual PURPLE messages. This can be accomplished in one of two ways: by paper and pencil using Figures 6 through 9 and following the example in the "Deciphering a PURPLE Message" section of this paper, or with the software simulator.

ACKNOWLEDGEMENTS

We are indebted to Kayoko Weierud for her help in translating Japanese text which, like so many examples of bureaucratic communications, did not always make sense. The entire staff of the National Archives, College Park, MD, was universally helpful, but Barry Zerby, Deborah Edge and Louis Smith deserve special mention.

REFERENCES

1. Budiansky, S. 2000. *Battle of Wits*. New York NY: The Free Press.

2. Deavours, C. and L. Kruh. 1985. *Machine Cryptography and Modern Cryptanalysis*. Norwood MA: Artech House.

3. Kelley, S. J. 2001. *Big Machines: Cipher Machines of World War II*. Laguna Hills CA: Aegean Park Press.

4. Kruh, L. 1978. A Catalog of Historical Interest. *Cryptologia*. 2(4) 341-344.

5. Rowlett, F. 1998. *The Story of Magic*. Laguna Hills CA: Aegean Park Press.

6. Sinkov, A. 1996. *Elementary Cryptanalysis*. Washington DC. The Mathematical Association of America.

7. U. S. Navy. *R. I. P. 77 - Instructions for M-5 (Orange Diplomatic "B" Machine) (Purple)*. NARA. RG38. (Records of the Chief of Naval Operations), Naval Security Group, Crane Indiana, Radio Intelligence Publications (RIPs) 1924-1945, RIP No. 77.

APPENDIX A:
PURPLE MOTION INDICATORS[22]

SECRET
R.I.P. 77

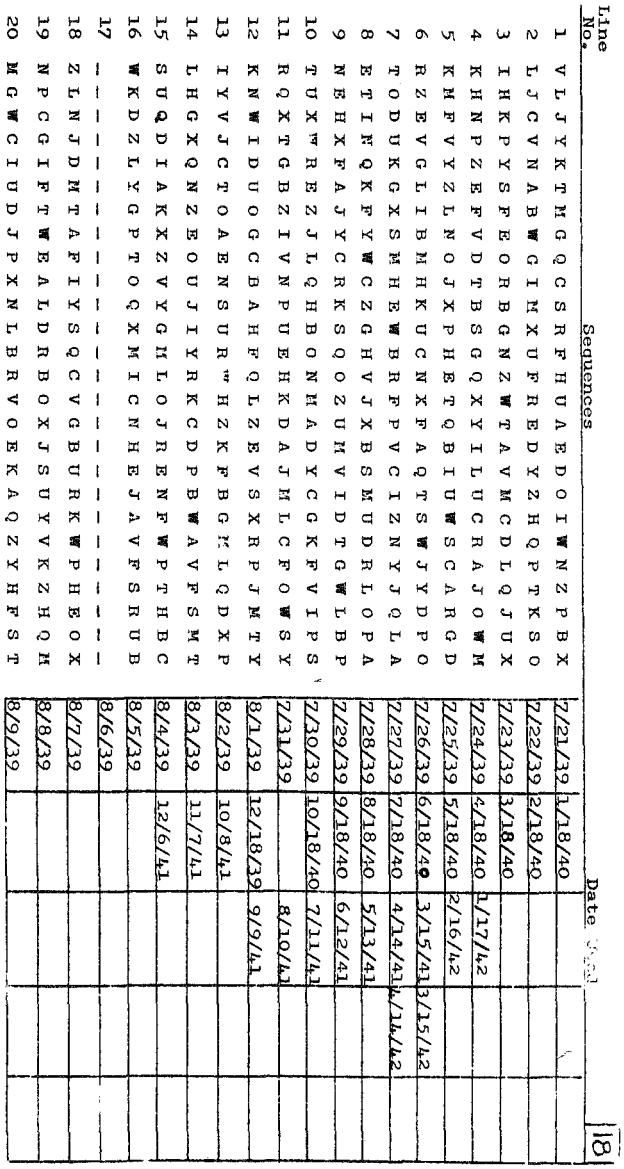
CHANGE NO. 1
1 April 1941

INDICATORS
STARTING POINTS
(cont'd)

Key	Six Wheel	1	2	3	20-Wheel Motion	Key	Six Wheel	1	2	3	20-Wheel Motion
13579	3	24	8	25	3-2-1	57139	19	20	6	4	2-3-1
13795	15	21	1	11	1-2-3	57391	25	10	25	20	2-1-3
13957	21	13	14	7	3-2-1	57913	23	16	19	13	1-2-3
15397	18	25	4	14	1-2-3	59173	6	19	15	10	3-2-1
15739	9	1	24	6	2-3-1	59317	20	18	2	17	3-2-1
15973	12	12	12	1	2-3-1	59731	14	5	3	3	1-3-2
17359	22	17	10	9	2-1-3	71395	8	3	11	8	1-2-3
17593	5	11	9	22	3-1-2	71539	5	6	2	23	3-2-1
17935	9	8	18	7	1-2-3	71953	6	1	22	4	2-1-3
19375	18	19	13	3	1-3-2	73159	23	5	20	1	2-3-1
19537	15	12	24	25	3-1-2	73591	19	18	19	20	3-2-1
19753	3	10	23	14	1-3-2	73915	24	3	4	19	2-3-1
31597	17	21	25	12	2-3-1	75193	10	20	9	5	3-2-1
31759	21	2	21	2	1-2-3	75319	25	24	10	24	1-3-2
31975	16	7	17	16	2-3-1	75931	2	4	16	18	3-2-1
35179	8	15	7	17	3-2-1	79135	22	23	12	9	1-2-3
35791	14	16	5	11	2-1-3	79351	12	9	1	13	2-3-1
35917	4	11	14	15	3-1-2	79513	11	25	3	6	1-3-2
37195	20	14	8	8	2-3-1	91357	1	22	6	10	1-3-2
37519	13	13	15	21	3-2-1	91573	7	17	11	22	3-1-2
37951	6	1	5	19	1-3-2	91735	8	23	1	9	2-3-1
39157	16	25	9	14	2-3-1	93175	25	7	6	12	1-3-2
39571	4	9	16	22	2-3-1	93517	10	13	18	15	3-2-1
39715	17	5	12	6	3-2-1	93751	20	3	25	8	1-2-3
51379	14	2	8	18	2-1-3	95137	19	4	7	21	2-3-1
51793	3	20	19	2	1-3-2	95371	12	21	20	11	3-1-2
51937	21	14	3	16	1-3-2	95713	11	6	22	20	3-1-2
53197	23	24	15	25	2-1-3	97153	18	22	11	17	2-3-1
53719	15	8	13	7	1-3-2	97315	22	10	4	24	3-1-2
53971	9	11	23	23	3-1-2	97531	24	12	2	10	1-3-2

[22][7, p. A2-A-3]

APPENDIX B:
PAGE 18 OF THE PURPLE ALPHABETS[23]



Line No.	Sequences	Date	18
1	V L J Y K T M G Q C S R F H U A E D O I W N Z P B X	7/21/39	1/18/40
2	L J C V N A B W G I M X U F R E D Y Z H Q P T K S O	7/22/39	2/18/40
3	I H K P Y S F E O R B G N Z W T A V M C D L Q J U X	7/23/39	3/18/40
4	K H N P Z E F V D T B S G Q X Y I L U C R A J O W M	7/24/39	4/18/40 1/17/42
5	K M F V Y Z L N O J X P H E T Q B I U W S C A R G D	7/25/39	5/18/40 2/16/42
6	R Z E V G L I B M H K U C N X F A Q T S W J Y D P O	7/26/39	6/18/40 3/15/42 3/15/42
7	T O D U K G X S M H E W B R F P V C I Z N Y J Q L A	7/27/39	7/18/40 4/14/42 4/14/42
8	E T I F N Q K F Y W C Z G H V J X B S M U D R L O P A	7/28/39	8/18/40 5/13/42
9	N E H X F A J Y C R K S Q O Z U M V I D T G W L B P	7/29/39	9/18/40 6/12/42
10	T U X "R E Z J L Q H B O N M A D Y C G K F V I P S	7/30/39	10/18/40 7/11/42
11	R Q X T G B Z I V N P U E H K D A J M L C F O W S Y	7/31/39	8/10/42
12	K N W I D U O G C B A H F Q L Z E V S X R P J M T Y	8/1/39	12/18/39 9/9/42
13	I Y V J C T O A E N S U R " H Z K F B G M L Q D X P	8/2/39	10/8/42
14	L H G X Q N Z E O U J I Y R K C D P B W A V F S M T	8/3/39	11/7/42
15	S U Q D I A K X Z V Y G M L O J R E N F W P T H B C	8/4/39	12/6/42
16	W K D Z L Y G P T O Q X M I C N H E J A V F S R U B	8/5/39	
17	- -	8/6/39	
18	Z L N J D M T A F I Y S Q C V G B U R K W P H E O X	8/7/39	
19	N P C G I F T W E A L D R B O X J S U Y V K Z H Q M	8/8/39	
20	M G W C I U D J P X N L B R V O E K A Q Z Y H F S T	8/9/39	

[23] [7, p. A3-B19]

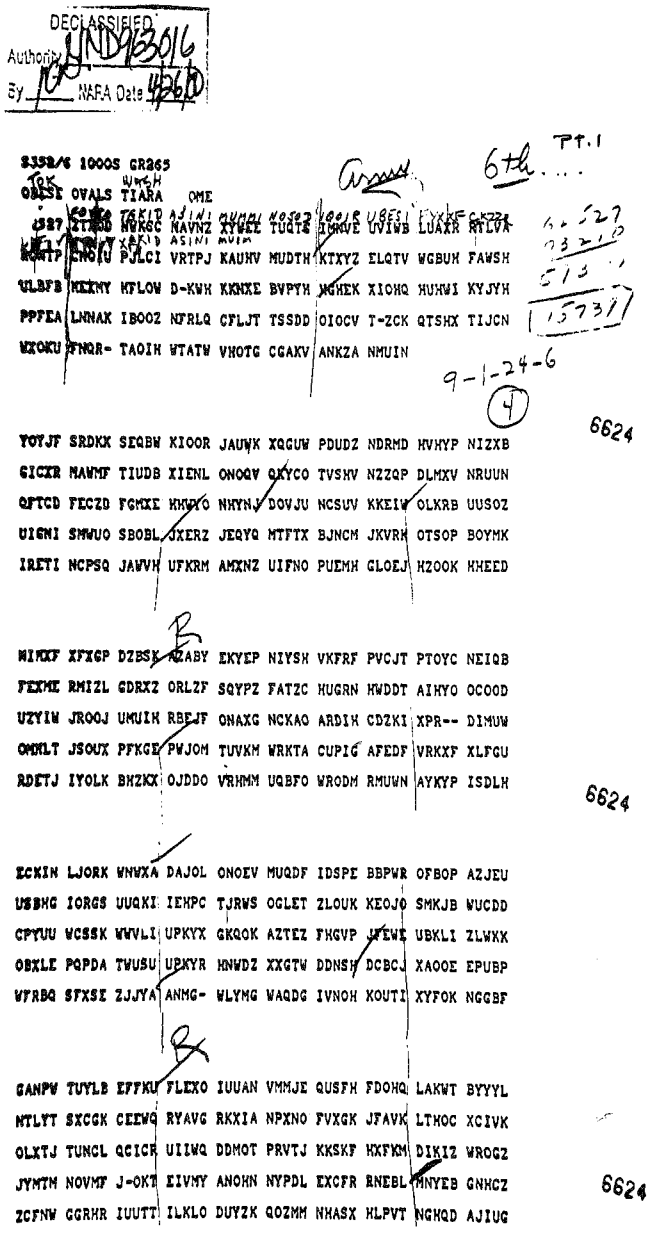
APPENDIX C:
PUNCTUATION CODES USED IN PURPLE MESSAGES[24]

	C	L	V	X	A	I	U	E	O
FA	REP. IND.	4TH REP.	木	(e)	1	BEGIN CHINESE	7	BEGIN FOREIGN SPELL	INSERT MARK OR WORD
FI	BEGIN KANA	1ST REP.	5TH REP.	1	(d)	2	...CODE TEXT		END FOREIGN SPELL
FE	↔	END KANA	2ND REP.	6TH REP.	□	(c)	3		
FO	GOKU-HI	KANCHO FUGO FORMATE KAI	(	3RD REP.	7TH REP.	ハ	(b)	4	
FY	# DAI	SIKU	BUGAI-HI	)	END CHINESE	L	=	(a)	5
CF	•	´ ACUTE ACCENT	¨ UMLAUT	# DAI	—				
LF	SUB ¶	, COMMA	— SPACE	%					
VF	~ TILDE	BLANK	:						
XF	# DAI	%	NEW CLAUSE IND.	;					

	CF	LF	VF	XF
C	¶		OPEN QUOTES	a
L	TITLE UNDER-LINE FOR SUB ¶	CAPITAL INITIAL LETTER	b	CLOSE QUOTES
V	' APOS-TROPHY	C	- HYPHEN	
X	d	ABBR. POINT	SUB-SUB ¶	SPACE OR DASH

[24]NARA. RG457. NSA Historical Cryptographic Collection. *Purple Machine Abbreviations and Garble Chart*. NR. 2839, Box 952.

APPENDIX D:
CIPHERTEXT OF PART 1 OF THE 14-PART MESSAGE[25]



[25]NARA. RG457. NSA Historical Cryptographic Collection. *Copies Of Messages 901, 902, 907, and 910 From Tokyo To Wash On 6 And 7 Dec 41*. NR. 1815, Box 738.

APPENDIX E: CIPHERTEXT OF A PURPLE MESSAGE FROM TOKYO TO BERLIN, DATED JULY 27, 1942[26]

12/203/29 DGY DE JUM 13705KCS S4 R3

SSS 11 SCDE TOKIO 2 28 27 455S JG 1/50

KOSHI BERLIN

DAIQU 68759 NUQOR ZEVEB VJWNO HNJKR NHDYU TSQJO PEXML THYII

IJIEN DKTYT RZBIC PMEMF TSOMC GXEPS KSDUL CSFTU XGQGN RULXY

UJRSS QIUES GDHOC TVHQA RMLKH MCJRZ FHGDS HTHVK EEKYG AKVKZ

EEXCC EZWYL EUKPN NLCKH OBTNJ WEVRD TMZNX VJUFO MHWNJ SHHBS

RJKJO RHUTL HYHAJ JLZUW ICETI SOGRG ABTUZ SLDEZ

P2/11 KOS 50WDS

STIOI OYTNZ UJUKB JQCFY TNUCG YIBQF AZKNG JOEWO ZWDPV NHMLU

QTNLY JQYJD GLCSO OZVKJ EDCST ZAVDC JCBRZ QDKRT SHAGY MGHSL

VIVPP UGXNE GNEWT KRNCM RAEIH PCNOH KUCBY FJJOE VRJKK THFYW

AZRTR TRFNM VPMDC LCRFB TJDAI PHNNK IZONW BPHOL HXJWT EGIMV

MOVCJ DSVLV WUYGN EXUNB VZEFE JDSAH UDHBZ VPGEJ UYMJW KGJKK

P3/11 KS 50WDS

PICFT MUAWZ POCPU XHVQA YCWEB RIKOU HSHVW AEWWT JILGB NKXVL

WNEOZ TTSGY HKUNS AGHKV KUZMV RINTN OBKKY RYBNP XTIOM HEVYC

GMOQF RUULM BUARD YMIWS OKMTU VUBGD ITGEW IDNHQ BGKSO EZFKG

ZIZUJ HBSEY WPMVY JSISX SCLUR DDQHB VVGSH GOAEK ADDPE VMMNS

WJOPT CQDYZ FBDOZ XNXXE FEQMU PNXSK ZRAYA AHJGW KTCVL ECIQK

P4/11 KOS 50

HIMEP AASBH VEVRW BKPGY OYLAA EGGVH MTNBC BFMKC NFNHX FJGHW

RKOCE KULUP UGRUV OYOWD JVNIZ -VNTQ MRLSD MEEYE XMIDY TOJOL

UUUST BEBUG SBQDT XDDKL CCMYL CQGEZ IMFIK ABMYI FPHKC FGJNM

NWFAE DLNBY EMETD GSIYN FRBUH KJGDA VDKLW PWRGG UM-WH XNXMU

KGG-C LIZSP CZWCS RBBYX TNRKC UUUUR BYOGD JJHZM JEOAA MZSWH

P5/11 KOS 28WDS

JFYZK RMVNW JPVGS RHCHL CMKSY VZTRE KMKLI UQHLK TYPYR WOLVF

EUKJP FNLUS FQTGZ UNIGV M2MWI HJCPF MAEDG GQBSB FCIFL IHETI

DFCYV GJCWS IDBBQ RUBQW EHDST TUNSZ AABZK

TOGO

0643GMT/27

[26]NARA. RG457. NSA Historical Cryptographic Collection. *JAPANESE CIPHER AND TRANSLATIONS OF C. I. MESSAGES ON USE OF TSU CODE*. NR. 2050, Box 774

BIOGRAPHICAL SKETCHES

Wes Freeman retired recently after a career as an analog applications engineer in the semiconductor industry. When not touring on his bicycle, he studies WW2 cipher machines.

Geoff Sullivan is a computer programmer and electronics engineer working on the design of scientific instruments. His main interest in cryptography is the computer simulation and computer cryptanalysis of historic cipher machines.

Frode Weierud is employed by the European Organization for Particle Physics (CERN) in Geneva. He works as a programmer in one of the equipment groups. Cryptography has been his main interest for more than 35 years. His cryptological research is focused on cipher machines and cryptanalytical techniques.