1.0: Introduction:

60 years before this page was originally written, the cryptographers of Hut 8 (Naval Enigma) at Bletchley Park (BP) perfected "Banburismus", a unique statistical attack that would work against German Navy Enigma messages even if they had no cribs available (i.e no guesses as to the exact plaintext). Even when cribs could be derived, Banburismus was still useful in being able to reduce the number of possible wheel orders to be run on the bombes. The choice of three out of eight wheels in Naval Enigma meant that there were 336 possible wheel-orders to check, and any clues that could reduce this count would greatly speed up the solutions.

The purpose of this page is to rectify the fact that most references to Banburismus just mention it in passing. Few people ever knew how it worked, nowadays almost nobody knows how it worked and fewer again could claim to have used the procedure since WWII. Here you will find the entire technique described from the theory up, culminating with a worked example on some English language messages enciphered with German Navy Enigma procedures.

1.1: "Banburismus":

The name "Banburismus" is actually used for two things. It is the overall name for all the methods described here which, taken together, are used to break a Naval Enigma ciphertext. These include "decibanning", "dummyismus" and "scritchmus" and we will meet them all below.

Fundamentally though, Banburismus refers to an optical aid to comparing two messages against each other, looking for characters which match in both. It is done by placing the messages (represented by holes punched in sheets of card) on top of each other over a light-box and counting places where the light shines through. Light can only shine through where two holes are in the same place - representing matching characters at that position in the messages. (It seems that there was a variation on this theme where the two messages would be put on top of each other over a dark table-top. Where the holes lined up the dark table-top would be visible and would contrast with the light-coloured card.) The card sheets were printed in Banbury (a town in central England, not far from Bletchley Park) and were known as "banburies" - and the technique was "banburismus".

The technique of Banburismus has been so thoroughly lost that no-one these days even knows exactly what a 'banbury' looked like. Tony Sale, in an article to the BPARK internet mailing list stated that when he prepared props for the Channel 4 series "Station X", he made sheets with letters printed on a half-inch pitch horizontally and vertically. Surviving Hut 8 veterans commented to him that they "looked about right".

Some evidence says that Tony made his sheets a bit too big:

- [KAHN1996] claims (page 141 of the paperback) that banburies were 10 inches high and various widths from 2 foot to 5 foot. Since there are 26 letters vertically then this statement precludes Tony's half-inch letter spacing.
- BP are known (from [WELCH1997] page 220) to have used third-inch letter spacing on their clones of Zygalski sheets which they presented to the Poles in early 1940. Such a choice of measurement aroused curiosity from the metric-thinking Poles! The choice might have been prompted by the fact that normal commercial ring-binder punches are typically

quarter-inch diameter, and would be compatible. Not just that, but they would be available even in wartime Britain.

Having made the third-inch choice once, it seems very likely that BP's next punched card data-processing system would use the same ideas. Additionally, if banburies were third-inch spaced, there would be space for guide-numbers (mentioned in [ALEX1946], para. 21) along the top edge and sensible margins at the top and bottom.

With third-inch column-spacing, Kahn's comment about 5 foot wide sheets would tie in with sheets capable of taking 180 character messages or thereabouts. However, [ALEX1946] para. 21 comments that banburies ranged from 100 column to 250 column which rather suggests that the column-spacing was actually quarter-inch.

There is no reason why banburies had to use the same spacing horizontally and vertically. However, quarter-inch spacing for the columns would make it possible for two punch holes to overlap if indeed commercial ring-binder punches were used for that purpose. It may be that slight overlap of holes didn't matter - the sheets only had to survive for a day or so once punched anyway.

1.2: Enigma Procedures in General:

The Enigma machine generates a large number of simple substitution ciphers, one for each letter of the message being enciphered. The setup for the machine can be considered to consist of two phases, The "inner setting" involves opening the machine's covers, selecting the three rotors to be used, setting the ring on each and fitting them into the machine. It was deemed an "officers only" procedure. The inner settings were changed whenever necessary. Up to the mid 1930's they were only changed every few months, but during WWII most users of Enigma were expected to do it daily. The exception was in the navy, who changed inner settings approximately every two days for no obvious reason.

The "outer setting" consisted of wiring up the plugboard - a rather less tricky procedure. Typically ten pairs of letters were "steckered" in this way, the remaining six being left unplugged.

The only part of the ciphering mechanism visible to the operator are the three or four letters showing the current positions of the rotors. At any given setting of the rotors, the machine can be said to generate an "alphabet" i.e a mapping of plaintext to ciphertext. Normally, the whole alphabet of the machine is not seen, because once one letter has been enciphered, the rotors step onwards to the next setting where a whole new alphabet would be available.

Of course - any two Enigma machines with the same inner and outer setting will generate the same alphabets at the same rotor-settings - this is what makes it possible to cipher and decipher messages.

About 17000 rotor-settings are available on the three-rotor machines, and nearly 440000 on the four-rotor U-boat and Abwehr machines. The procedures for sending messages on Enigma machines attempt to cause as many as possible of the available rotor-settings to be used so as to minimise the chance of characters in multiple messages using overlapping rotor-settings (such an occurrence was known as messages being "in depth") which might in turn give away the contents of messages, or worse, give away the whole inner setting of the machine and therefore the contents of all messages that day. So starting all messages with "AAA" as the rotor-setting is **not** permissible!

Messages are enciphered starting from different initial rotor-settings therefore, but so that decryption of Enigma-enciphered messages is possible by the intended recipient, this chosen 'message-setting' must somehow be transmitted along with the message itself. Obviously it can't be sent in plaintext or it would still be easy for attackers to make use of "depths" wherever they might be found. Typically, Enigma messages start with a short group of letters known as an 'indicator'. An indicator is the message-setting enciphered starting at a different rotor setting (known as the grundstellung). This has only shifted the problem however - now the user has to communicate the grundstellung somehow! Different services of the Wehrmacht did this in different ways:

1.2.1: Army and Luftwaffe Enigma Procedure:

For each message sent, the Army and Luftwaffe expected the cipher clerk to choose a unique three-letter group as its grundstellung and just send it in plaintext at the start of that message. If done right, then cryptographically speaking this is a very good system, however in practice various operator errors gave BP's cryptanalysts many ways to attack those messages. No more will be said about them though, as they are well covered in other literature, and Banburismus cannot work against them anyway for reasons that will be stated below:

1.2.2: Naval Enigma Procedure:

For each of their cipher networks, the Navy printed the grundstellung to be used on a given date in the same key-list that contained the inner and outer settings for the machine. This meant that the grundstellung didn't have to be transmitted, it was there in the recipient's key-list.

The key to Banburismus (and the reason that it won't work on Lufwaffe and Army messages) was that this use of a constant grundstellung meant that the indicators of all naval Enigma messages in the same cipher network were effectively 3-letter messages "in depth". The fact that they were in depth was not seen as a weakness by the Naval cipher authorities because they took steps to make sure that the message-settings were very random (they were picked from a printed list in a book called the Kenngruppenbuch). Additionally, the indicators themselves were not transmitted directly, they were super-enciphered with a bigram-replacement hand-cipher.

It's probably best to illustrate the procedure with a cut-down example. Let's assume that the alphabet only contains the letters A, B and C. Every cipher operator in the German Navy (regardless of which cipher network they used) will have been issued with a copy of:

- The Kenngruppenbuch (a.k.a "K-Book").
- The Zuteilungsliste (or "usage table") for the K-Book.
- Nine sets of bigram substitution tables.

Cipher operators whose ships sent Enigma messages (not all of them did) will additionally have been issued with a keylist of inner and outer settings for the machine and the grundstellungs. These settings-lists were unique to the particular cipher network to be worked.

1.2.2.1: The Kenngruppenbuch:

For our cut-down-alphabet example, a simplified K-Book looks like this:

Section A

1	2	3	4	5
CAB	BCC	ABA	CCA	ABC
BAC	AAB	BBC	BCA	BBB
ABB	BBA	CBA	BAA	ACB
BAB	AAA	CAC	CBC	
AAC	CBB	CCB	ACA	
BCB	CAA	ACC	CCC	

Page 1

Section B

AAA 2 B 2 C 1	**AB**A 3 B 1 C 5	**AC**A 4 B 5 C 3
BAA 4 B 1 C 1	**BB**A 2 B 5 C 3	**BC**A 4 B 1 C 2
CAA 2 B 1 C 3	**CB**A 3 B 2 C 4	**CC**A 4 B 3 C 4

Page 1

The real K-Book was of course much bigger than this and contained 26^3 three-letter groups. Section A was arranged as 732 columns of 24 three-letter groups and one final column (number 733) containing just 8 groups. There were 20 of these columns printed per page, thus Section A contained 37 pages in all. Section B was printed with 26 look-up columns per double-page spread. The column headed "AAA" would be on the top left of a left page, with the column headed "ABA" next to it, running horizontally to "AGA". The column headed "AHA" would be below "AAA" on the left page, and the bottom right column of that page would be "ANA". The sequence started again at the top left of the right-hand page with columns "AOA" through to to "ATA" printed above the columns "AUA" through to "AZA". The following two-page spread in the book would be "BAA" through to BZA". Section B thus took up 52 pages of the K-Book, meaning that there were 89 active pages in total. (More of a "Heft" than a "Buch" then...)

There is an illustration of a very small part of Section A of a real K-book in the Appendix to [KAHN1996], page 288.

The Navy didn't reprint the Kenngruppenbuch any more often than it considered neccessary. It was certainly changed at least twice between 1939 and 1945 but historians seem unsure of exactly when or how many times it changed.

1.2.2.2: The Zuteilungsliste (Usage List):

As we shall see shortly, part of the enciphered preamble of a German Navy message told the recipient the cipher network and the category of an incoming message. In Hut 8, Banburismus was only directed against "Home Waters" messages, whose cipher network was known as "Dolphin" in Hut 8, but as "Hydra" by the Germans.

Again, for our worked example, we shall use a cut-down Zuteilungsliste for our cut-down K-book as follows:

Cipher:	K-book columns:	
	Cipher M:	Hand cipher:
Potato	All: 1, 3	General: 1 Officer: 3
Carrot	All: 2	
Beetroot		General: 4 Officer: 5

"Cipher M" (above) is terminology used in the real Zuteilungsliste to refer to Enigma machine ciphers. Some cipher networks (like "Potato" above) specify a backup hand cipher to let those ships communicate even if the Enigma machine is destroyed or otherwise unusable.

The non-Enigma ciphers used the Kenngruppenbuch columns to distinguish officer-only messages from normal ones. Evidently the "Beetroot" cipher network is one such. The Enigma ciphers handled officer-only messages by double-enciphering them.

There is an illustration of a small part of a real Zuteilungsliste in the Appendix to [KAHN1996], page 289.

1.2.2.3: The Doppelbuchstabentauschtafel (Bigram Substitution Table):

As a final level of complexity, the message preamble was enciphered with a bigram lookup table. Again, for our example, there's a cut-down example below:

AA CB	**B**A CC	**C**A CA
B BC	B AC	B AA
C BB	C AB	C BA

The German Navy would typically have a set of 9 of these tables in use at any one time, each known by the letters of the alphabet "A" through "J" (no letter "I"). One of these tables was specified to be used on each day. It was only a matter of time before some or all of these substitution tables were "pinched" or reverse-engineered. Without them (as will be seen) Banburismus cannot be used.

An important characteristic of the real Bigram tables was that they were reciprocal. The was intended to make the system less prone to human error, but also helped BP's reverse engineering efforts. The reciprocal nature of the tables is reproduced in the cut-down example above. The eagle-eyed reader will have noticed that "CA" maps to "CA" in the example too. This is *not* a feature of the genuine tables, and is caused here by having a cut-down alphabet with an odd number of letters! Obviously, the real tables had 26^2 entries (which is even) and there was thus no need for an identity mapping.

There is an illustration of a small part of a bigram table labelled "E" in the Appendix to [KAHN1996], page 288. However, this table is evidently not reciprocal, and is labelled "Verschlüsseln" ("Encipherment"), implying that a separate table was needed for deciphering.

Either that table belongs to some other bigram substitution cipher system, or dates from some point where the Navy used non-reciprocal tables for the Kenngruppen. Historical evidence seems to point to the tables being reciprocal for at least the 1941 - 1943 period that saw the use of Banburismus.

There is an illustration of a real reciprocal bigram table (table "B" from the "Fluss" set) in [ERSK1992] and [BAUER2007], p62.

1.2.3: A worked example of German Navy Procedure:

Imagine that we are enciphering a message to be sent in the "Potato" cipher. Indeed, for most German ships, there would be only one cipher that they could use, they wouldn't know the Enigma machine's inner and outer settings nor the Grundstellung for any other systems.

The Zuteilungsliste says that we are to indicate our use of "Potato" with a three- letter group from columns 1 or 3 from Section A of the K-book. We shall choose "CAC" from column 3 and cross it off with a pencil.

Next, we need an "indicator group" for our message. We can choose any three-letter group from the K-book for this. We decide on "ABB" from column 1 and cross it off too.

We must now use the Bigram table of the day to compose the first two 4-letter groups of our message. We write down "CAC" and "ABB" as follows on the message-sheet:

	C	A	C
A	B	B	

We fill in the spare positions with "randomly chosen letters":

B	C	A	C
A	B	B	B

Now we look up the vertical pairs starting from the left and write the bigram-table mappings horizontally. In this case, the first pair is "BA" and its mapping is "CC". The next pair "CB" maps to "AA", followed by "AB" which maps to "BC" and finally "CB" which maps to "AA".

Our first two four-letter transmission groups are therefore "CCAA" and "BCAA". Next, we must start enciphering the message itself. We set the Enigma machine to the Grundstellung specified for "Potato" for this day in the keylist ("ABA" for instance). We type in the indicator that we chose from the K-book (which was "ABB") and note the result. Let's say that the result was "BCA". That's the message-setting!

We now roll the rotors of the machine around to ""BCA" and start entering the message itself, and we write the results down as four-letter groups to follow the two four-letter groups that we've already generated by hand with the bigram tables. When all the groups are ready, then they are transmitted in morse code on the right frequency band for "Potato" messages for that particular time of day.

1.2.4: Meanwhile at BP:

Hut 8 would get a copy of this message from an intercept station but would not be able to proceed with deciphering it unless they had access to at least part of the Bigram Substitution table for the day. If they did have such a table and if that table was sufficiently complete that they could decipher the "indicator" group, then the message could be considered for Banburismus. If at least 200 such messages could be isolated from the day's catch, then Banburismus could be started. Analysis of letter-repeats between pairs of enciphered messages could (with care) reveal the distances between their message-settings. Careful dovetailing of this information could even reveal the plaintext of some of the third letters of the indicators (which were all enciphered at a rotor-setting of grundstellung+2). Such knowledge could form a bombe menu, and from that the key could be obtained. Additionally, certain properties of the distances between letters on the right-hand wheel could eliminate some of the wheels or even reveal exactly which wheel it was, and that cut the number of wheel orders that had to be run on the bombes - thus speeding up solutions.

1.3: First Success: Foss's Day:

On 26th April 1940, British forces captured the armed trawler "Polares" and (more importantly) captured some Naval Enigma keylists, matching plaintexts and ciphertexts and other material. This allowed Hut 8 to break Naval Enigma for the period April 22nd through April 27th by about mid June that year according to Joan Murray in [HINS1993], page 113. (This event is sometimes referred-to as "The Narvik Pinch" ([ALEX1946], [MAHON1947]), though it apparently occurred at Alesund - nowhere near Narvik. The boat wasn't called Narvik either!) BP's week of successfully reading Naval Enigma yielded cribs for use in the future, but more critically allowed BP to reverse-engineer parts of the bigram substitution tables. That in turn allowed Banburismus to be tried for the first time on selected days where the true indicators of a reasonable number of messages could be deciphered with those partial tables. The first **successful** use of Banburismus was against the key for 8th May 1940, though it took until November 1940 for the solution to be found! Hugh Foss (later head of Japanese Naval Section in Hut 4) was the man credited with the feat, and 8th May was henceforth known as "Foss's Day" in recognition of it.

The capture of the "Krebs" in early 1941 allowed a similar reconstruction attack on the bigram substitution tables then in force. These were known as "Bach" and were different from the ones reconstructed from the "Narvik Pinch". A set of tables typically stayed in force for upwards of 5 months. The "Bach" tables were in force from 1st July 1940 to 14th June 1941.

Actual bigram tables were captured from U110 on 9th May 1941 (see [KAHN1996], chapter 13). However, it would seem that these would have been the same as those reconstructed after the "Krebs Pinch". The bigram tables changed to a set called "Fluss" on June 15th 1941, but the U110 had been at sea since early March and would have been due home in late May/early June. [SALE2000b] claims that the capture of "München" (on May 7th 1941) provided the keys for all of June 1941 and this is confirmed by [ALEX1946] para 31. Alexander doesn't mention "München" by name but we know from Hinsley and others that it was that ship.

From this pinch, the next set of substitution-tables ("Fluss") were reverse-engineered as the intercepts for the latter half of June 1941 came in. This is a perfect example of a fatal silly mistake; one month's key-list being allowed to span the changeover between bigram tables.

2.0: Repeats, Overlaps and Messages "in depth":

Banburismus depends on the fact that if two messages in German (or any other natural language) are compared letter-for-letter, the chance of finding matches between the letters is higher than it would have been if the messages had been just random letters.

Consider the following two messages (as with real Enigma messages, the spaces have been removed):

```
HereisthefirstmessageofapairNothingspecialjustordinaryEnglishtext
BelowitanotherAgainyoucanseethatitismerelyarandomexamplemessage
 -    -                -        -  - -             -       -  -
```

Matching characters (known as 'repeats' at BP) are underscored. We get nine single-character repeats in an overlap of 62. Had the messages been random, we would have expected one repeat approximately every 26 characters - so maybe we'd get two or three in an overlap like this.

This property is not changed if both messages are enciphered at the same setting of an Enigma machine:

```
GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQSSPZXARIXEABQIRUCKHGWUEBPF
UXOLKADJZLMWVBTSPSBHXIZGWJAUNOHDXPXEWSHMZWULSAJZFNEQGCWRLZFWLCB
 -    -                -        -  - -             -       -  -
```

Each letter of the message has been changed into a different one, but obviously if the letters had originally been the same, their encipherments will also be the same.

This property of an Enigma machine or any other polyalphabetic cipher was first documented by American cryptographer William Friedman in the 1920's, and apparently rediscovered independently by the Polish cipher bureau in the 1930's as they made their breaks into Enigma.

2.2: Indicators.

If we look at our two messages when not enciphered at the same setting of the Enigma machine, then comparing them letter for letter yields results more like what we'd expect from random text.

```
GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQSSPZXARIXEABQIRUCKHGWUEBPF
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVUQILBJUABNLKMKDJMENUNQ
                    -                                       -
```

The indicators for the above two messages happen to be VFG and VFX respectively.

2.3: Searching for "Evidence".

The first task to be done in Banburismus then is that of comparing pairs of messages like those above at all possible offsets to see if the pattern of repeats looks promising. This in turn (if we're lucky) might show us how far apart the rotor-settings of the Enigma machine must have been when they were enciphered.

BP adopted a shorthand for writing down the repeats that appeared. For instance, they might write "$9^{xx}/56$" meaning a total of 9 repeats including two bigrams (noted by the 'X's), in an overlap of 56 characters. More extreme cases might be "$20^{3xx}/161$" meaning 20 repeats including a trigram and two bigrams in an overlap of 161 characters.

Those two messages with indicators VFG and VFX (above), compared at all offsets from -25 to +25 give repeats as follows:

VFG=VFX-25:	**1/40**
VFG=VFX-24:	**4/41**
VFG=VFX-23:	**1/42**
VFG=VFX-22:	**2/43**
VFG=VFX-19:	**1/46**
VFG=VFX-18:	**2/47**
VFG=VFX-17:	**4/48**
VFG=VFX-16:	**1/49**
VFG=VFX-15:	**2^{X}/50**
VFG=VFX-14:	**4^{X}/51**
VFG=VFX-13:	**6/52**
VFG=VFX-11:	**3/54**
VFG=VFX-9:	**9^{XX}/56**
VFG=VFX-8:	**3^{3}/57**
VFG=VFX-7:	**2/58**
VFG=VFX-5:	**3/60**
VFG=VFX-4:	**2/61**
VFG=VFX-3:	**3/62**
VFG=VFX-2:	**3/63**
VFG=VFX-1:	**4^{X}/63**
VFG=VFX	**impossible!**
VFG=VFX+1:	**2/63**
VFG=VFX+2:	**1/63**
VFG=VFX+3:	**3/62**
VFG=VFX+4:	**4/61**
VFG=VFX+5:	**1/60**
VFG=VFX+6:	**3/59**
VFG=VFX+7:	**3/58**
VFG=VFX+9:	**2/56**
VFG=VFX+10:	**3/55**
VFG=VFX+11:	**3/54**
VFG=VFX+12:	**1/53**
VFG=VFX+14:	**1/51**
VFG=VFX+15:	**1/50**
VFG=VFX+16:	**1/49**
VFG=VFX+18:	**2/47**
VFG=VFX+19:	**3/46**

VFG=VFX+20: **1/45**
VFG=VFX+21: **1/44**
VFG=VFX+22: **2/43**
VFG=VFX+23: **1/42**
VFG=VFX+24: **1/41**

Before we even started comparing those messages above we could tell that their message-settings can be no more than 25 letters apart one way or the other because the first two characters of both indicators (VFG and VFX) are the same. We therefore know that the first two letters of the message-settings are the same, and that means that no mid-wheel or left-wheel turnover occurred between them. A turnover will happen at least every 26 letters. If a double-notched "Navy Wheel" is used as the right-wheel, a turnover will happen every 13 letters. We can also tell that there's no point in checking with an offset of 0, because for them to be in depth with an offset of zero, their indicators would have to be the same!

The situation at offset VFG = VFX-9: a repeat of "**9^{XX}/56**" would seem auspicious:

```
GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQSSPZXARIXEABQIRUCKHGWUEBPF
         YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVUQILBJUABNLKMKDJMENUNQ
                       - --  -   -          -  -   --
```

It turns out that two bigrams and seven other matches in an overlap of 56 is almost 5:1 in favour of the theory that we're looking at two (English) messages "in depth" with the initial Enigma machine settings for those messages being nine letters apart. The next best obvious repeat seen at offset VFG = VFX-8 features a trigram (but nothing else):

```
GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQSSPZXARIXEABQIRUCKHGWUEBPF
        YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVUQILBJUABNLKMKDJMENUNQ
                 ---
```

This situation is simple enough that we might guess that VFG=VFX-9:(giving **9^{XX}/56**) beats VFG=VFX-8: (giving **3^{3}/57**) but we can't be sure. In fact this assumption is true, the former case is 5:1 in favour, the latter is barely 2:1 in favour. But to resolve these issues in more complex cases, BP needed some way of rewriting the repeats as "scores" where higher scores were more favourable than lower scores.

Luckily, a little-known 200 year old branch of mathematics known as **Bayesian Statistics** deals with this very concept.

3.0: Bayesian Statistics.

A reader not interested in the gory details can skip this section, but should note the points made in the highlighted boxes.

3.1: Some theory:

Bayesian statistics was developed by the 18th century English presbyterian minister, Rev. Thomas Bayes (1702 - 1761). His key paper was published posthumously in 1763 and revealed Bayes's Theorem.

Put simply, this theorem gives the probability of a event being true based on the result of some "prior knowledge" that you may have as modified by one or more tests that you can make. It derives this from the probability of the test being true assuming the event is true.

So it is a way of turning probabilities around. In the case of Banburismus, we want to test for the case that two enciphered messages are "in depth" given a count of the numbers of characters, bigrams, trigrams (etc) are the same in the two messages.

Bayes's Theorum

$$P(J \mid W) = \frac{P(W \mid J)P(J)}{P(W \mid J)P(J) + P(W \mid {\sim}J)P({\sim}J)}$$

The notation P(J) means 'the probability that J is true', and P(J | W) means 'the probability of J being true given that W is true'. The notation P(~J) means the probability that J is not true. Obviously P(~J) = 1 - P(J).

I shall not prove this theorem here. If you are interested in the proof, check out [BERRY1996], pages 147-148.

One interesting feature of Bayesian statisticians is that they almost always refer to probabilities in terms of bookmaker's odds rather than the more commonly seen probabilities expressed at numbers from 0 to 1 (or 0% to 100%). There is a very good reason for this.

The odds of something happening is the ratio of the probability that it will happen to the probability that it will not. In other words:

$$O(J) = \frac{P(J)}{P({\sim}J)}$$

If we rewrite Bayes's theorem in terms of odds we find a considerable simplification (for the full derivation, see [BERRY1996], page 148-149):

Bayes's Theorum expressed in Odds

$$\frac{P(J \mid W)}{P({\sim}J \mid W)} = \frac{P(J)\ P(W \mid J)}{P({\sim}J)\ P(W \mid {\sim}J)}$$

or...

$$O(J \mid W) = O(J) * \frac{P(W \mid J)}{P(W \mid \sim J)}$$

The notation O(J) means 'the odds in favour of J being true' and is known as the **prior odds** . The notation O(J | W) means 'the odds in favour of J being true given that W is true' and this is known as the **posterior odds** . The term P(W | J)/P(W | ~J) is known as the **Bayes Factor** .

It should be fairly obvious that the posterior odds from one test can be used as the prior odds on a subsequent test. This means that an overall posterior odds can easily be found from the results of several tests applied to a situation. The only important rule to observe is that all tests must be independent of each other, and all must be independent of the original prior odds.

Bayesian statisticians (especially those with no computers like the banburists of Hut 8) often make life easier for themselves by working with logarithms of odds. This means that the multiplications of the prior odds and Bayes factors become simple additions. Also, conveniently, odds of 1:1 (evens) is represented by a log-odds of zero. Negative log-odds represent "odds against" and positive log-odds represent "odds in favour".

4.0: Scoring Charts:

4.1: Decibans and "HubDubs":

The direct calculation of log-odds of Bayes factors for every possible combination of "evidence" (i.e monogram repeats, bigrams, trigrams etc in every possible overlap length) would an impossible task today, let alone on the mechanical calculators of the WWII era. The obvious solution is for someone to compile tables (BP always referred to them as "charts") on which a given repeat can be looked-up and the relevant score just be written down. Tables of Logarithms, Sines, Cosines etc had been compiled in this way for hundreds of years - it was a natural way to handle such a problem in those days.

[ALEX1946] paragraph 31 notes that when charts were first compiled, they were tabulated in "decibans" to the nearest 0.1. A "deciban" is obviously a tenth of a "Ban" and a "Ban" was the log10 of the Bayes Factor (the name "Ban" being a shortening of Banbury). Later they moved to half-decibans rounded to the nearest integer. It would appear from [HINS1993] page 158, that this latter improvement was due to mathematician Jack Good. Working in whole numbers was deemed easier and less error-prone and Good reckoned possibly 50% of the time spent scoring repeats was saved. It seems though that even after the switch to half-decibans (known as hubdubs, written hdB), the procedure of turning observed repeats into scores was still known as "decibanning".

Pre-compiled charts of scores notwithstanding, there are a lot of variables to take into consideration (numbers of monograms, bigrams etc in various possible overlap lengths). Being brought up in an era of table-lookup however, BP's mathematicians knew a trick or two for simplifying things.

4.2: Score Charts at BP:

A glimpse of the actual layout of BP's charts can be deduced from [ALEX1946], para. 25 (but beware of a couple of clumsy corrections to the text). The description here is also slightly confused by issues to do with "dummyismus" (the handling of dummy messages). However, the layout of the score chart is revealed in that the procedure for scoring a repeat of "$11^{3x}/171$" is given - in other words a repeat consisting of 11 matching characters consisting of a trigram, a bigram (and by inference 6 monograms) all in an overlap length of 171.

It seems like a rather clunky nomenclature, but it is in fact an optimum shorthand for use with their charts. The charts consist of hdB scores tabulated horizontally for all lengths of overlap and vertically for all counts of monogram matches. In the above example, you'd look for column "171" and row "11", but then move down one 'bonus' row for each 'X' (bigram) noted and down an extra 4 'bonus' rows for every '3' (trigram) noted. So actually the score you'd want would be the one on row "16".

This system of applying "bonuses" to a base score in order to handle bigrams and trigrams in addition to the monograms is very convenient. **It is also a crude hack which just so happens to work well enough to be (presumably) worth the slight errors that slip in. You can't extend this convenience to tetragrams and beyond.** (It isn't obvious at this stage why not - all will be explained in section 4.5 below where the concept of "loss" is introduced.)

4.2.1: Handling Tetragram (and higher order) Repeats.

The scores due to any tetragram (and higher order) repeats have to be looked-up separately and added into whatever the score charts indicate for the low-order repeats. For instance a repeat of "$21^{643x}/181$" would be dealt-with by looking up the scores for a hexagram and a tetragram and adding their total to the score for the remainder of the repeat (i.e "$11^{3x}/171$" which was illustrated in the previous section).

4.2.2: Message Categories.

Originally, Banburismus was done with tetragram repeats always treated as equally likely amongst messages, and thus allocated a single fixed score. However, BP soon discovered that traffic analysis of the messages allowed them to refine this and account for the fact that certain messages would be more likely to yield tetragram repeats (see [ALEX1946], paras. 22 and 29). The reason for this is stated to be because naval Enigma messages had to have numbers spelt out in full. Many German numbers are four-characters in length (eins, zwei, drei, vier, funf, seqs, acht, neun). Actually, "zwei" is often rendered "zwo" as is still common practice when reading numbers over phone lines today, but the principle still holds.

Traffic analysis allowed BP to guess which types of messages were likely to contain numerals (requests for N litres of fuel or reports of number of torpedoes or other ammo expended for instance). Each message was allocated a "category" (a number from 1 to 20, written in roman numerals).

Once categorised, the score for a tetra appearing between two messages could be looked up on a table. It is claimed in [ALEX1946], para. 29 that a score for a tetra repeat was not even purely dependent on what category messages produced it, but even down to where in the messages it appeared! BP allocated 13 possible 'locations of interest' for tetras to start: positions 1 to 10 of the message (location 1 through 10), from 11 to 30 (location 11), in the middle (location 12), and within 30 from the end (location 13).

An example of handling tetragrams is given in [ALEX1946], para 26. It seems to contain the only clue as to the layout and values on BP's charts, however all we get is that a tetra in the middle of a category XI message scores +19 and that a tetra in the middle of a category XIV message scores +23. The total score for that tetra is therefore +42.

Tetra	Message Category																			
Location	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
3	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
4	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
5	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
6	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
7	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
8	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
9	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
10	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
11	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
12	?	?	?	?	?	?	?	?	?	?	19	?	?	23	?	?	?	?	?	?
13	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

We can never reproduce BP's tetra charts based on this information. The best we can do is shown above. It is interesting however to see the level of attention to detail which was lavished on Banburismus as a whole.

4.3: Duplicating BP's Scoring Charts:

The layout of the main scoring charts is revealed (in section 4.2 above) to depend only on monogram count (M) and overlap length (N). Given this, we can easily print our own charts once we've worked out how to calculate the Bayes Factors involved. Each Bayes Factor is (from section 3.1 above) the probability for M characters matching **and** N-M characters not matching in superimposed German texts divided by the same thing for superimposed random texts. We know that the probability of characters matching in random text is 1/26, and (obviously) the probability of them not matching is 25/26. We will represent the probability of characters matching in naval Enigma as "P" and therefore the probability of them not matching is (1-P).

The Bayes Factor for M monograms in an overlap of N is therefore:

$$\text{B.F.} = \frac{P^{M}*(1 - P)^{N-M}}{(1.0/26.0)^{M}*(25.0/26.0)^{N-M}}$$

....and of course for charts in hdB we would want to print 20 times the logarithms of these values (rounded to the nearest integer) for all M and N likely to happen in reality.

From [ALEX1946] para. 6 we learn that the probability of single characters matching in German naval messages was "about 1/17". A chart of the Bayes Factors (as in the formula above) with P set to 1/17 would be fine for the case where nothing diminishes the probability of characters matching. Actually, there are two key reasons why this is not very likely.

4.4: Complications - "Dummyismus" and "Distance":

The first cause of diminished probability is due to the chance that any given message is a dummy - or is a short German text that changes into gibberish padding some way through. The skill of handling these situations was known as "dummyismus" which we'll look at later (section 11). The result of the procedure however was the allocation of a "loss" score due to the chance of the messages being dummy.

The second way for a message to be allocated a "loss" is down to "distance". This is down to the rather non-obvious fact that if a repeat is seen at some number of characters separation from another, there is a chance that the repeat counts for nothing due to the fact that the middle wheel of the machine would have done a "carry" in that interval and would invalidate the assumptions given by the indicators.

An example makes it clearer. Suppose you have two messages with indicators JFQ and JFE, and you have a repeat of "$21^{43x}/200$" favouring JFQ = JFE+25. These messages are supposedly 25 characters apart, and if for a minute we pretend that only wheels I thru V exist, then there's obviously only a 1/26 probability that the middle wheel hadn't experienced a "carry" in this distance. So there's no point in looking up this repeat on a chart compiled with the idea that two characters in German texts will match with a probability of 1/17 . We need to take into account a "loss" of -28 hdB (i.e 20*log(1/26)).

(Referring to the losses in terms of hdB would seem to be the right thing to do because in the case of losses incurred from both "distance" and dummyismus, the losses can be added up and the correct chart consulted. [ALEX1946] doesn't make it very clear.)

Classical statistics theory says that if there is an overall probability L (ranging from 0 to 1) that a given repeat is valid, then instead of calculating charts of scores based on a probability 1/17 of characters matching, we must use a probability P, where:

$$P = \frac{L}{17} + \frac{(1 - L)}{26}$$

this works for all L, but actually we would be interested only in those values of L given by:

$$L = 10^{(t/20)}$$

where t are negative integers (given in hdB). Any calculations of loss that we perform will give negative integer results.

It would seem logical to propose that BP would refer to a score chart which tabulated the Bayes Factors with P set to 1/17 as a "zero chart" because it would represent the situation when total "loss" (L) is 0hdB. References to "1 charts" and "3 charts" will be discussed below in section 5.4.

4.5: Complications - the breakdown of the "bonus" system of scoring:

A consequence of the "bonus" system of scoring is that the score for (say) a trigram will be found on a "zero chart" if you look up the repeat "**3+B/3**", (where **B** is the bonus for a trigram - i.e 4). In other words:

$$\text{Trigram Score} = \frac{(1/17)^{7} * (1 - 1/17)^{3-7}}{(1.0/26.0)^{7} * (25.0/26.0)^{3-7}}$$

Now if we compare what we get out of a chart compiled for a loss of 't' hdb against applying the same loss directly to the Trigram Score, they'll be different. In other words:

$$\frac{P^{7} * (1 - P)^{3-7}}{(1.0/26.0)^{7} * (25.0/26.0)^{3-7}}$$

$$\text{where } P = \frac{10^{(t/20)}}{17} + \frac{(1 - 10^{(t/20)})}{26}$$

(the alleged score for a trigram from a chart compiled for a loss of 't') is not the same as:

$$\frac{10^{(t/20)} * (1/17)^{7} * (1 - 1/17)^{3-7}}{(1.0/26.0)^{7} * (25.0/26.0)^{3-7}} + 1 - 10^{(t/20)}$$

(which is the true score for a trigram compensated for a loss of 't' hdb).

5.0: End-Wheel Comparisons.

The so-called "End Wheel Comparison" is where scores are evaluated for repeats between pairs of messages where both the first and second letters of the indicators are the same. This was the assumed case in the discussion about 'distance' in section 4.3 above.

Before we can try to evaluate scores for end wheel comparisons, we'll have to construct a distance chart for that situation:

5.1: The Distance Chart:

We want to know the probability that the middle wheel does **not** turn over in a stretch of N characters. Let there be a chance 'S' that the end-wheel is a single-notched type (and therefore a chance 1-S of it being a double-notched "navy wheel"). If it's a single-notched wheel then there's a (26-N)/26 chance of no turnover, and if it's a "navy wheel" then there's a (13-N)/13 chance of no turnover. The overall probability of no turnover is:

```
    S*(26 - N)      (1 - S)*(13 - N)
L = ----------  +  ---------------
        26               13

... for 0 <= N <= 13 , otherwise:

    S*(26 - N)
L = ----------
        26

... for 13 < N <= 25
```

We are now in a position to recreate BP's 'distance chart' merely by repeating the above calculation for all N from 0 to 25. We don't print the probabilities 'L' directly, we print the hubdub equivalents (i.e. 20*log(L)):

```
Dist:     0   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
        ------------------------------------------------------------------------
------------------------------
Score:    0  -1  -1  -2  -2  -3  -4  -4  -5  -6  -7  -8 -10 -11 -12 -13 -14 -15 -
16 -17 -18 -20 -22 -24 -28 -34
```

Assuming an even distribution of wheel choice, then 'S' would be 5/8. However, it is claimed (in [SALE2000b]) that the inner settings of navy Enigma machines always included at least one wheel VI, VII or VIII. If so, this would change the value of the constant 'S' from 5/8 (0.625) to 150/276 (0.5435).

The primary effect of this would be to make losses slightly worse when N is large. This is what we might expect - the chance of there being a "Navy Wheel" in the right hand position must increase if rules say that at least one navy wheel must be present. Navy wheels cause turnovers every 13 letters, which worsens the chances of an uninterrupted run of letters longer than N=13.

There seems to be no clear knowledge of whether or not Hut 8 assumed there would always be at least one "Navy Wheel" present. Ralph Erskine (in private correspondence) agrees that it was usually the case, but cites a few isolated dates (early in the war) where the keylist didn't obey the rule.

Whatever the truth, the mathematics of the distance tables remains as stated above. It may never be known which value of 'S' was assumed by Hut 8. For the purposes of this document, it will be assumed that a key always included at least one "Navy Wheel".

5.2: Proof of the pudding:

As a demonstration of all this, we can work out the score for a repeat of "$7^{xx}/32$" representing TYQ = TYB+5. The 'distance' involved here is 5 letters, and the distance chart (above) tells us that it corresponds to a 'loss' of -3hdB. We'll need a chart compiled for a loss of -3hdB in order to evaluate the repeat. Here is a portion of such a reconstructed chart:

Number	Overlap Length											
of repeats	29	30	31	32	33	34	35	36	37	38	39	40
0	-4	-4	-4	-4	-4	-4	-5	-5	-5	-5	-5	-5
1	-1	-1	-1	-1	-1	-2	-2	-2	-2	-2	-2	-2
2	1	1	1	1	1	1	1	1	0	0	0	0
3	4	4	4	4	4	4	4	3	3	3	3	3
4	7	7	7	7	7	7	6	6	6	6	6	6
5	10	10	10	10	10	10	9	9	9	9	9	9
6	13	13	13	13	13	12	12	12	12	12	12	12
7	16	16	16	16	15	15	15	15	15	15	15	15
8	19	19	19	18	18	18	18	18	18	18	18	17
9	22	22	21	[21]	21	21	21	21	21	21	20	20
10	25	25	24	24	24	24	24	24	24	23	23	23
11	28	27	27	27	27	27	27	27	27	26	26	26
12	30	30	30	30	30	30	30	30	29	29	29	29
13	33	33	33	33	33	33	33	32	32	32	32	32
14	36	36	36	36	36	36	35	35	35	35	35	35

We first look up column '32' and row '7' giving us a score of 16. But we need to take into account the two bigrams, for which the rule is to drop down an extra row for each bigram (drop down four rows for each trigram). This gets us to a score of +21.

This very example is mentioned in [ALEX1946], para. 7, and the result quoted there is actually a score of +22. More on this discrepancy later....

5.3: A more complicated scenario:

The example in [ALEX1946], para. 25 takes into account losses due both to distance and to dummyismus. (We will just take Alexander's dummyismus figures as-is for now). Basically, we're trying to score a repeat between ASL = ASJ+5 where we have "0/30" before the 'blue line' and "$11^{3x}/171$" after it. Dummyismus says that before the blue line the messages are considered genuine, so no loss due to dummyismus, but there is a loss (not mentioned by Alexander) of -3hdB for distance. So the repeat "0/30" is looked up on a chart compiled for a loss of -3hdB (it just so happens that this is therefore the same chart as was consulted for the example in section 5.2 above). The relevant extract of the chart is as follows:

```
  Number  |                   Overlap Length
of repeats|  29   30   31   32   33   34   35   36   37   38   39   40
          +------------------------------------------------------------
    0     |  -4  [-4]  -4   -4   -4   -4   -5   -5   -5   -5   -5   -5
    1     |  -1   -1   -1   -1   -1   -2   -2   -2   -2   -2   -2   -2
    2     |   1    1    1    1    1    1    1    1    0    0    0    0
    3     |   4    4    4    4    4    4    4    3    3    3    3    3
    4     |   7    7    7    7    7    7    6    6    6    6    6    6
----------+------------------------------------------------------------
    5     |  10   10   10   10   10   10    9    9    9    9    9    9
    6     |  13   13   13   13   13   12   12   12   12   12   12   12
    7     |  16   16   16   16   15   15   15   15   15   15   15   15
    8     |  19   19   19   18   18   18   18   18   18   18   18   17
    9     |  22   22   21   21   21   21   21   21   21   21   20   20
----------+------------------------------------------------------------
   10     |  25   25   24   24   24   24   24   24   24   23   23   23
   11     |  28   27   27   27   27   27   27   27   27   26   26   26
   12     |  30   30   30   30   30   30   30   30   29   29   29   29
   13     |  33   33   33   33   33   33   33   32   32   32   32   32
   14     |  36   36   36   36   36   36   35   35   35   35   35   35
```

Alexander claims a score of -5 for this repeat.

Now for the repeat "**11^{3x}/171**". Alexander comments that dummyismus allocates a loss of -4 to these two messages beyond the blue line. There is already a loss of -3 in force due to distance, so the repeat needs to be looked up on a chart compiled for a loss of -7hdB:

```
  Number  |                       Overlap Length
of repeats| 167  168  169  170  171  172  173  174  175  176  177  178
          +------------------------------------------------------------
   10     |   5    5    5    5    5    5    4    4    4    4    4    4
   11     |   7    7    7    7    7    6    6    6    6    6    6    6
   12     |   9    9    9    9    8    8    8    8    8    8    8    8
   13     |  11   11   11   11   10   10   10   10   10   10   10   10
   14     |  13   13   13   12   12   12   12   12   12   12   12   12
----------+------------------------------------------------------------
   15     |  15   15   14   14   14   14   14   14   14   14   14   14
   16     |  17   16   16   16  [16]  16   16   16   16   16   16   16
   17     |  18   18   18   18   18   18   18   18   18   18   18   18
   18     |  20   20   20   20   20   20   20   20   20   20   20   19
   19     |  22   22   22   22   22   22   22   22   22   22   21   21
----------+------------------------------------------------------------
   20     |  24   24   24   24   24   24   24   24   24   23   23   23
   21     |  26   26   26   26   26   26   26   26   26   25   25   25
   22     |  28   28   28   28   28   28   28   28   27   27   27   27
   23     |  30   30   30   30   30   30   30   29   29   29   29   29
   24     |  32   32   32   32   32   32   31   31   31   31   31   31
```

We get a bonus of 1 for the bigram, and a bonus of 4 for the trigram in the repeat and thus look up 16/171 and find a score of +16 (Alexander claims +19).

We now add up these two scores to get an overall score for the repeat of +12. This differs a fair bit from Alexander's result (he claimed +14).

5.4: Score discrepancies and a couple of curiosities.

The reconstructed tables are evidently close, but not quite right. Errors of ±1 compared with Alexander's figures could easily be due to either a different technique for integer roundoff (especially in the roundoff used to tabulate the distance chart), or maybe a different way of generating the scoring charts in the first place. Certainly it is impossible for BP to have calculated each individual score in the table from scratch as was done here.

There is a possible clue in that [ALEX1946] refers to the charts used in the example above as a "1" chart and a "3" chart respectively. This isn't explained at all. It is **possible** that maybe the naming convention for these charts dated back to before the switch to hubdubs for scoring purposes, the "1" chart being compiled for a loss of one deciban, the "3" chart for a loss of three decibans.

If that was true, it means that the scoring charts consulted in the examples of sections 5.2 and 5.3 above were the wrong ones. A "1" chart would correspond to a loss of -2hdB, and a "3" chart to a loss of -6hdB. If we repeat the scoring using such charts we get a score of +24 (too high) for the example of section 5.2, and scores of -4 and +18 (both out by just 1, but their sum is correct at +14) for section 5.3.

What we need here is the testimony of a surviving "Big Room Girl" or one of the original mathematicians. How **did** BP get from the total loss (in hdB) to the knowledge of which chart to consult? Given the care BP took elsewhere to tune the last hdB's worth of accuracy from their scoring system, it seems ridiculous that they would treat losses as carelessly as to merely round them off to the nearest deciban (as seen above) - this leads to ±2 or ±3 changes in score.

An extra problem is caused by the account of scoring in [CLIFF1943] in which it is made very clear that three and only three scoring tables were available, known as the "1", "2" and "3" charts. [CLIFF1943] elaborates very slightly on [ALEX1946] in stating:

- The "1" chart was used for situations of virtually no "loss".
- The "2" chart was used for situations where there was a "75% chance of being correct" (i.e about 2.5hdB loss).
- The "3" chart was for "50% chance of being correct" (i.e about 6hdB loss).

Specifically, [CLIFF1943] states that there were no other charts for higher amounts of loss.

Evidently, though the mathematics of Banburismus are clear (and will be demonstrated below), the exact procedures of Hut 8 are not completely known. It could well be that higher amounts of loss were dealt with by looking up a basic score on the "3" charts and applying an additional correction factor by hand.

5.5: The Bayesian Prior.

In order to compile the so-called "fit lists" (in which most promising repeats were tabulated), the Banburists had to calculate for each repeat, the actual odds for that repeat being correct. This involves taking the overall score for the repeat and correcting for the Bayesian Prior Odds (see section 3).

The Bayesian Prior Odds are the odds of a situation being correct before any tests have been performed on it. In Banburismus, when working with just the end-wheel (as we are in all the examples above), the prior odds of a given repeat being the right one are 50:1 against (there are 25 possible offsets to consider on either side of zero, but zero itself isn't a contender if the indicators are different).

Log odds of 50:1 against is obviously -34hdB. So 34 needs to be subtracted from the score for a repeat before we can evaluate the odds that it is correct.

We can prove this (somewhat) by looking at [ALEX1946], para. 13. Here is an example of a fit-list, showing some example repeats together with their odds of being right.

Amongst them is STK = STN+7, a repeat of "**$23^{33}/256$**" for which Alexander claims odds of 15:1 on. That's log odds of +23.5hdB, which must mean that the actual score for the repeat was +57.5hdB by the time the prior odds have been accounted-for. The reconstructed charts for a loss of -4hdB (corresponding to a distance of 7) give a score of +52.

This is obviously not right, but follows the trend that the modern charts seem to under-score w.r.t BP's originals. However, there is a mystery in that odds of 15:1 as claimed by Alexander don't correspond to an integer number of hdB as might be expected. Not only that, but the two nearest integers values of hdB come out as +23hdB -> 14.1:1 on (we'd expect this to be rounded to 14:1 on), and +24hdB -> 15.8:1 on (which we'd expect to be rounded to 16:1 on). A claim of 15:1 on seems rather strange. It just has to be the case that Alexander took his example from the days when the charts were still in decibans to one decimal place.

5.6: Deciban Sheets.

Alexander often refers to Deciban Sheets as being used to record the scores awarded for all ±25 possible offsets checked with the banburies when handling an end-wheel comparison. There is a fine example in [ALEX1946] para. 16.

At first glance it seems overkill actually to score **every** repeat for all ±25 possible offsets checked with the banburies. Alexander offers no direct reason for it, apart from the comment at the end of [ALEX1946] para 26 in which a score of +18 gleaned from a mid-wheel comparison is "entered accordingly on the S and T deciban sheets". By this, it would seem sensible to assume that the +18 is added to whatever score was found on the deciban sheets at that particular offset between S and T.

After all, the whole principle of the Bayesian Technique is that scores from different tests all pointing to the same fact can be added together to get a composite (and hopefully more accurate) score for that fact being true. Actually, by extending the argument a bit it would be sensible to conclude that **all** end-wheel comparisons that are made between S and T get entered together on the **same** deciban sheet together with any contributions from mid-wheel comparisons that feature S and T. They get added together in the manner of matrix-addition for matrices each consisting of a single

50-element row. This would explain why all 50 scores are evaluated for end-wheel comparisons - the matrix arithmetic needs it thus.

It is easy to spot the best score between pairs of letters after the matrix arithmetic. For instance, when we try Banburismus for ourselves in Section 8, we will find the following data assembled on the V and W deciban sheet:

```
V = W+?
-25-24-23-22-21-20-19-18-17-16-15-14-13-12-11-10 -9 -8 -7 -6 -5 -4 -3 -2 -
1     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
```

```
  0  0  0 -1  0 -1  0  2  0 -4  0  6 -3 -5  3  0 -8  6 -6  6-12 -8 -
8 23  0  W  3-12 -8  7 -3 -6 -6 -2  1 -7  6  3 -2 -1 -2 -4  0  2  0 -
1  1  0  1  0  0    ZO[V|W]
  0  0  0 -1  2  0 -2 -2  0  1 -1 -4 -1  2 -2  9 -2 -1 -7 17  2  2 -5  4-24  W -
6 -6  4 -3 -4  2  4  5 -1 -9 -7 -2  0 -5  2 -1 -1  0  0 -
1  0  1  0  0  0    GM[V|W]
  0  0  0 -2  2 -1 -2  0  0 -3 -1  2 -4 -3  1  9-10  5-13 23-10 -6-13 27-24  W -
3-18 -4  4 -7 -4 -2  3  0-16 -1  1 -2 -6  0 -5 -1  2  0 -
2  1  1  1  0  0    [27] V=W-2     (2.2:1 against)
```

From this we can see that the ZO$^{V}/_{W}$ repeat was indicating a best score of +23 for V=W-2, and the GM$^{V}/_{W}$ repeat indicated a best score of +17 for V=W-6. A naiive interpretation might be that the +23 score beats the +17, and the true distance should be V=W-2. When the whole 50-element matrices are added (bottom row) we find that the score for V=W-2 has improved to +27 and that the score for V=W-6 has increased to +23. So V=W-2 is indeed the distance to believe (it happens to be correct too!). A score of +27 is still worse than evens (the Bayesian Prior is -34), but such a score comes in handy when disambiguating which of three possible end-wheel alphabets is correct later in the procedure.

More on all of this in section 8....

6.0: The Middle-Wheel alphabet.

The principles of scoring repeats between messages where both the second and third letter of the indicators differ is exactly the same as detailed above in section 5 (End-Wheel Comparisons), but the distance charts need recompiling to allow for both the middle and end wheels, and when converting scores to odds, there's the small matter of the Bayesian Prior being 1300:1 against (i.e -62hdB rather than -34hdB).

There is also one major operational difference. It isn't practical to lay two Banburies over each other on the light-table and write down all possible repeats. There are theoretically ±650 possible offsets to consider. See below (section 6.2) for comments about how BP partially managed to automate this process.

6.1: A Distance chart for the middle and end wheels.

This chart is a bit more complex than was the case for the end-wheel-only distance-chart. We are interested in the probability that the left-hand wheel has not turned over in a given stretch of letters.

Of course, whether or not it has experienced such a turnover depends on what sort of wheels are used in the middle and end positions.

There are four possibilities: NN, AN, NA and AA where 'A' refers to a single- notched "Army wheel" and 'N' refers to a double-notched "Navy wheel". With always at least one navy wheel present, the chance of AA wheels is 60/276 with the left wheel experiencing a 'carry' every 650 (i.e 25*26) letters. The chance of AN wheels is 90/276 with the 'carry' every 325 (i.e 25*13) letters. The chance of NA wheels is also 90/276 but with the carry every 312 (i.e 12*26) letters, and finally the chance of NN wheels is 36/276 with the carry every 156 (i.e 12*13) letters.

The reason why the left-wheel carry is different in the 'NA' and 'AN' situations is down to the rather odd behaviour of the carry mechanism. In brief, when the middle wheel moves to where its notch engages the left-wheel, it moves again on the next letter, taking the left-wheel with it. So the distance between carry-points of the middle-wheel is one fewer than it would have been had that wheel been on the right. Confused? Grab a real Enigma machine and watch the sequence for yourself.

Proceeding as we did in the case of the end-wheel only, the overall probability of no turnover on the left wheel is:

```
     60   (650 - N)         90   (325 - N)         90   (312 - N)         36   (156
- N)
L = --- * --------    +   --- * --------    +   --- * --------    +   --- * ----
----
    276      650          276      325          276      312          276
156

... for 0 <= N < 156 , otherwise:

     60   (650 - N)         90   (325 - N)         90   (312 - N)
L = --- * --------    +   --- * --------    +   --- * --------
    276      650          276      325          276      312

... for 156 <= N < 312 , otherwise:

     60   (650 - N)         90   (325 - N)
L = --- * --------    +   --- * --------
    276      650          276      325

... for 312 <= N < 325 , otherwise:

     60   (650 - N)
L = --- * --------
    276      650

... for 325 <= N < 650
```

As above, we can now recreate BP's 'distance chart' merely by repeating the above calculation for all N from 0 to 649. Again, we don't print the probabilities 'L' directly, we print the hubdub equivalents:

```
      |                                               Letters
      |  0   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
------|-------------------------------------------------------------------------
-------------------------------
   0  |  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-1  -1  -1  -1  -1  -1  -1  -1
   1  | -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
-1  -1  -1  -1  -1  -1  -2  -2
   2  | -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2  -2
-2  -2  -2  -2  -2  -2  -2  -2
   3  | -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3  -3
-3  -3  -3  -3  -3  -3  -3  -3
   4  | -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4  -4
-4  -4  -4  -4  -5  -5  -5  -5
   5  | -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -5  -6  -6
-6  -6  -6  -6  -6  -6  -6  -6
   6  | -6  -6  -6  -6  -6  -6  -6  -6  -6  -6  -6  -7  -7  -7  -7  -7  -7  -7
-7  -7  -7  -7  -7  -7  -7  -7
A  7  | -7  -7  -7  -7  -7  -7  -8  -8  -8  -8  -8  -8  -8  -8  -8  -8  -8  -8
-8  -8  -8  -8  -8  -8  -8  -8
l  8  | -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9  -9 -10 -
10 -10 -10 -10 -10 -10 -10 -10
p  9  |-10 -10 -10 -10 -10 -10 -11 -11 -11 -11 -11 -11 -11 -11 -11 -11 -11 -11 -
11 -11 -12 -12 -12 -12 -12 -12
h 10  |-12 -12 -12 -12 -12 -12 -13 -13 -13 -13 -13 -13 -13 -13 -13 -13 -13 -14 -
14 -14 -14 -14 -14 -14 -14 -14
a 11  |-15 -15 -15 -15 -15 -15 -15 -15 -15 -16 -16 -16 -16 -16 -16 -16 -16 -17 -
17 -17 -17 -17 -17 -18 -18 -18
b 12  |-18 -18 -18 -18 -18 -18 -19 -19 -19 -19 -19 -19 -19 -19 -19 -19 -19 -19 -
19 -19 -19 -19 -20 -20 -20 -20
e 13  |-20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -20 -
20 -20 -20 -20 -20 -20 -20 -20
t 14  |-20 -20 -20 -20 -21 -21 -21 -21 -21 -21 -21 -21 -21 -21 -21 -21 -21 -21 -
21 -21 -21 -21 -21 -21 -21 -21
s 15  |-21 -21 -21 -21 -21 -21 -21 -21 -21 -22 -22 -22 -22 -22 -22 -22 -22 -22 -
22 -22 -22 -22 -22 -22 -22 -22
  16  |-22 -22 -22 -22 -22 -22 -22 -22 -22 -22 -23 -23 -23 -23 -23 -23 -23 -23 -
23 -23 -23 -23 -23 -23 -23 -23
  17  |-23 -23 -23 -23 -23 -23 -23 -23 -23 -24 -24 -24 -24 -24 -24 -24 -24 -24 -
24 -24 -24 -24 -24 -24 -24 -24
  18  |-24 -24 -24 -24 -25 -25 -25 -25 -25 -25 -25 -25 -25 -25 -25 -25 -25 -25 -
25 -25 -25 -25 -25 -25 -26 -26
  19  |-26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -26 -27 -27 -27 -
27 -27 -27 -27 -27 -27 -27 -27
  20  |-27 -27 -27 -27 -28 -28 -28 -28 -28 -28 -28 -28 -28 -28 -28 -28 -28 -28 -
29 -29 -29 -29 -29 -29 -29 -29
  21  |-29 -29 -29 -29 -30 -30 -30 -30 -30 -30 -30 -30 -30 -30 -30 -31 -31 -31 -
31 -31 -31 -31 -31 -31 -31 -32
  22  |-32 -32 -32 -32 -32 -32 -32 -32 -33 -33 -33 -33 -33 -33 -33 -34 -34 -34 -
34 -34 -34 -34 -35 -35 -35 -35
  23  |-35 -35 -36 -36 -36 -36 -36 -36 -37 -37 -37 -37 -37 -38 -38 -38 -38 -39 -
39 -39 -39 -40 -40 -40 -41 -41
  24  |-41 -42 -42 -42 -43 -43 -43 -44 -44 -45 -45 -46 -47 -47 -48 -49 -50 -50 -
51 -53 -54 -56 -57 -60 -63 -70
```

6.1.1: Modulo-26 distances.

The layout of the distance table above is designed to suit BP's practice of writing distances greater than 25 characters in a modulo-26 format formatted "A.L", where 'A' is a count of alphabets, and 'L' is letters.

There is a really good reason for doing this. When handling the mid-wheel alphabet, the 'L' part of the offset relates directly to the motion of the end-wheel, and the 'A' part of the offset relates indirectly to the motion of the middle-wheel.

N.B: When dealing with distances, in order to convert the "alphabet count" to true motion of the middle-wheel we need to know the location of the turnover notch(es) on the end-wheel.

6.2: The Freebornery.

In order to deal with pairs of messages whose indicators only share the first letter, overlaps from -650 to +650 would (in theory) have to be checked with Banbury sheets. In practice this was limited to ±250 due to the naval Enigma procedures which ruled that no message could be more than 250 characters in length

Even so, this would have been impossibly time-consuming, but luckily, BP managed partially to automate this part of their codebreaking work and used Hollerith punched-card machinery to search for tetragram repeats between messages. For this work they used a central Data Processing facility run by a Mr. Frederick Freeborn. This was Hut 7/Block C.

Freeborn would produce the "tetra catalogue" which was a printout of all tetragram (and better) repeats found between pairs of messages whose indicators shared only the first letter.

6.2.1: How it was done.

[WHELA1944] reports that Hut 8 would mark up a ciphertext, giving it a message-number and splitting the ciphertext into segments each of five 5-letter cipher groups. Freeborn's card-punch operators would create a deck of master cards with each card as follows:

Card Column	Contents	Notes
1-3	Message number	
4-5	Card Number (i.e Segment number)	
6-7	Underlap	Last two cipher characters of the previous segment
8-32	The twenty-five ciphertext characters of this segment	
33-37	Overlap	The first five cipher characters of the next segment

[WHELA1944] notes that the exact column-usage of the cards was not necessarily exactly thus, but serves to illustrate the discussion. However, we do know that a few fields have been left out of the list above:

- We are told that columns 1-3 of a card was a "message number". This is backed up by [USN1943], but [USN1943] states that the indicator letters for the message were punched on the cards too. [ALEX1946], para 11 agrees and shows indicator letters present on the tetra catalogue. Doing so makes sense for the Freebornery - they can save a huge amount of work by separating messages with different first indicator letters. Hut 8 are not interested in knowing that a tetra exists between messages whose first indicator letters differ, and this is backed up by [USN1943] stating that message decks were manually sorted at punch-time into one of 26 "folders" according to the first indicator letter.

  However, [ALEX1946], para 11 contradicts this somewhat by showing the tetra catalogue displaying an entry for AJU = BYX - 1.23.

- There is no mention of any provision on the cards to keep track of where in a message the tetra on a given card "came from" (so to speak). Yet [ALEX1946], para 11 shows that the tetra catalogue did have that information printed on it in the modulo-26 notation similar to mentioned in section 6.1.1 above. There must have been a couple of columns on every card that would be sequentially punched on the message deck, presumably done by a tabulating punch after the message deck has been verified.

  There is a quirk in the usage of modulo-26 absolute letter-positions though: [ALEX1946], para 11 specifically states that an absolute letter-position of "4.8" represented 3*26 + 8 (i.e 86 letters into a message). This implies that the 'A' and 'L' counters ran from 1 to N, and therefore "1.1" would represent the first letter of a message. This is clearly backed up by [USN1943]. However, when two absolute positions were subtracted to give a "distance", the nomenclature reverts to the one detailed in 6.1.1 above. [ALEX1946], para 14 refers to distance of "0.17" to represent 17 characters distance, and [MAHON1947] backs this up by using quite a few distances with the 'A' part left blank (i.e zero).

  This "count from 1" behaviour appears to be confined to the numbering of the absolute positions of the tetras printed in the tetra catalogue. It is possible that the card-punches couldn't start from zero or simply that people were not used to numbering systems starting from zero in those days and that "first letter of first alphabet" was more intuitive to them.

Following [WHELA1944]'s claims, the master cards for the "VFG" message (used as example in 2.2 above) might look as follows:

```
04501  GXCYBGDSLVWBDJLKWIPEHVYGQZWDTH
04502GQZWDTHRQXIKEESQSSPZXARIXEABQIRU
04503EABQIRUCKHGWUEBPF
```

(We shall assume that Hut 8 numbered it as message 45 for the day. We shall also ignore for a moment the fact that the cards also held the indicator and counters for the absolute position within the message.)

Having created the master cards, these were now fed into the primary feed of a collating machine (IBM Type 77?) programmed to slip 24 blank cards after every master card. This deck was now put into a reproducing punch programmed to copy the master card's columns 1-5 to the 24 cards following, but to punch columns 7-37 into the blank card's columns 6-36 as follows:

```
04501  GXCYBGDSLVWBDJLKWIPEHVYGQZWDTH
```

```
04501 GXCYBGDSLVWBDJLKWIPEHVYGQZWDTH
04501GXCYBGDSLVWBDJLKWIPEHVYGQZWDTH
04501XCYBGDSLVWBDJLKWIPEHVYGQZWDTH
04501CYBGDSLVWBDJLKWIPEHVYGQZWDTH
04501YBGDSLVWBDJLKWIPEHVYGQZWDTH
04501BGDSLVWBDJLKWIPEHVYGQZWDTH
04501GDSLVWBDJLKWIPEHVYGQZWDTH
04501DSLVWBDJLKWIPEHVYGQZWDTH
04501SLVWBDJLKWIPEHVYGQZWDTH
04501LVWBDJLKWIPEHVYGQZWDTH
04501VWBDJLKWIPEHVYGQZWDTH
04501WBDJLKWIPEHVYGQZWDTH
04501BDJLKWIPEHVYGQZWDTH
04501DJLKWIPEHVYGQZWDTH
04501JLKWIPEHVYGQZWDTH
04501LKWIPEHVYGQZWDTH
04501KWIPEHVYGQZWDTH
04501WIPEHVYGQZWDTH
04501IPEHVYGQZWDTH
04501PEHVYGQZWDTH
04501EHVYGQZWDTH
04501HVYGQZWDTH
04501VYGQZWDTH
04501YGQZWDTH
04502GQZWDTHRQXIKEESQSSPZXARIXEABQIRU
04502QZWDTHRQXIKEESQSSPZXARIXEABQIRU
04502ZWDTHRQXIKEESQSSPZXARIXEABQIRU
04502WDTHRQXIKEESQSSPZXARIXEABQIRU
04502DTHRQXIKEESQSSPZXARIXEABQIRU
04502THRQXIKEESQSSPZXARIXEABQIRU
04502HRQXIKEESQSSPZXARIXEABQIRU
04502RQXIKEESQSSPZXARIXEABQIRU
04502QXIKEESQSSPZXARIXEABQIRU
04502XIKEESQSSPZXARIXEABQIRU
04502IKEESQSSPZXARIXEABQIRU
04502KEESQSSPZXARIXEABQIRU
04502EESQSSPZXARIXEABQIRU
04502ESQSSPZXARIXEABQIRU
04502SQSSPZXARIXEABQIRU
04502QSSPZXARIXEABQIRU
04502SSPZXARIXEABQIRU
04502SPZXARIXEABQIRU
04502PZXARIXEABQIRU
04502ZXARIXEABQIRU
04502XARIXEABQIRU
04502ARIXEABQIRU
04502RIXEABQIRU
04502IXEABQIRU
04502XEABQIRU
04503EABQIRUCKHGWUEBPF
04503ABQIRUCKHGWUEBPF
04503BQIRUCKHGWUEBPF
04503QIRUCKHGWUEBPF
04503IRUCKHGWUEBPF
04503RUCKHGWUEBPF
04503UCKHGWUEBPF
04503CKHGWUEBPF
04503KHGWUEBPF
04503HGWUEBPF
```

```
04503GWUEBPF
04503WUEBPF
```

For every message of N characters, N-4 punch-cards would be produced. A typical day's processing according to [WHELA1944] might result in 80,000 cards being produced! As soon as each message was punched up, its tetragrams were subjected to a "breakdown sort" on column 8, thus creating twenty-six growing piles of cards representing "A???" to "Z???" tetras. With the inclusion of the information about "folders" from [USN1943], that would seem to imply 26 "folders" of 26 growing piles of cards....

Sorting was done literally by the "bucket sort" algorithm - that's how sorting machines worked. However, alphabetic sorting on IBM equipment in the 1930s was a two-pass procedure because alphabetic symbols had two holes per column. First you'd sort a given column on rows 'Y', 'X' and '0', then take all the 'Y' cards and sort on rows '1' - '9', ditto with the 'X' cards and with the '0' cards. BP had quite a few sorting machines, and the second pass of alphabetic sorting would be split across three machines to save time.

6.2.2: The "Cut":

When Hut 8 called "cut!" then it was time to sort all the "breakdown" decks on columns 9 - 11 so that all matching tetragrams could be found. This would happen between 15:00hrs and "late into the evening" on days when the wheel order was changed. It couldn't happen before at least a reasonable amount of traffic had been punched up, and after Hut 8 had identified the right hand wheel for the day. There wasn't much point otherwise.

According to [USN4193], Freeborn would work through the cards folder by folder, sending the results to Hut 8 as each folder was finished.

6.2.3: Producing the Tetra Catalogue:

Freeborn's staff will have done the post breakdown sorting on columns 9, 10 and 11 as a multipass bucket sort on just the one deck. After all, the 80,000 cards comprising the day's job was already broken into 26 'folders' according to the first character of the message indicator. That's about 3100 cards per folder. The breakdown sort has already split each 'folder' into 26 further piles of about 120 cards each. You now sort the deck numerically first on column 11, then pick up all the bins in order (bin 0 at the top, bin 9 at the bottom) and re-sort on column 11's Y, X and 0 zones. If the bins are again picked up in order (bin Y at the top) then column 11 will now be in alphabetical order. The deck is then passed to another machine, set for column 10 sorting, and the same two passes are performed again. Finally it is passed to a third machine for two passes of column 9 sorting. At that point the deck is in alphabetical order of tetras, all with the first letter the same.

Simulations based on the 450cpm (cards per minute) IBM type 080 sorter indicate that it would take just over 3.5 minutes to perform this task. This machine had been available since 1925 and is a likely candidate for the job.

As the column 9 sorter finishes a deck, the next partially-done deck (representing a set of tetras with some other first letter) should be rolling off the column 10 sorter and yet another be coming off the column 11 sorter.

Each finished deck must now be scanned for repeating tetras which are obviously going to appear next to each other. [WHELA1944] states that this last stage in producing the tetra catalogue (before printing it) was to run the sorted cards through a collator (again, probably a Type 077) in order to spot these identical tetras. This collating machine runs at around 240cpm which forms a bottleneck that can easily be bypassed by use of two such machines.

Finally, according to Whelan, all repeat tetras would be printed onto a listing using an IBM 405 tabulator ready to be sent to Hut 8. The tabulator is slower than even the 077 collator at 150cpm, but its workload was comparatively low (it only had to print out the cards with matching tetras) and one machine would easily cope on its own.

Merely assuming that Freeborn could bring three type 080 sorters and two type 077 collators to bear on the job implies that the sorting of all 26 breakdown decks in one 'folder' could be done in around one and a half hours. This seems to be about four to five times slower than expected. An implication of the fact that the Freebornery bothered to do an initial breakdown sort at all is that they had enough equipment to process at least two breakdown decks in parallel. If they could actually process three decks at once, Hut 8 would be receiving a folder roughly every half-hour through the night. It's still a thirteen hour job to get the whole day's traffic through at that rate though. If Freeborn could sort four decks at once, it's nine and threequarter hours, if five decks at once, then it's seven and threequarter hours.

Information is scant about how fast the Freebornery really worked, but [USN1943] does mention a scenario where the first folders arrive in Hut 8 at about 17:00 and the middle-wheel alphabet is done by midnight. That's seven hours of course, but it doesn't mean that Hut 8 had to have absolutely all the folders to hand before they could get enough of the middle-wheel alphabet to compose a bombe menu.

The tetra catalogue merely points out where these high-order repeats were found, it would still require someone to set up the same overlap with Banbury sheets in order to evaluate the whole repeat. This meant that BP didn't normally consider it worth accounting for the slight risk that they might miss cases where the correct repeat contained only bigrams and trigrams. However there is a note in [ALEX1946] para. 20 to the effect that Freeborn could produce a "trigram catalogue" for use on days where insufficient data was coming out from the end-wheel comparisons and tetra catalogue.

6.3: Using the Tetra Catalogue:

The example in [ALEX1946], para. 26 evaluates a middle-wheel comparison and takes into account processing the tetra catalogue along with losses due both to distance and to dummyismus. (We will just take Alexander's dummyismus figures as-is for now). Basically, we're trying to score a repeat between CRS = CQT+2.3 giving "**$18^{4xx}/266$**" or maybe "**$19^{4xx}/266$**" (there's a glaring disagreement between lines 5 and 6 on page 106). Dummyismus says that the CRS message is considered 15% likely to be a dummy, and the CQT message has a 45% chance of being dummy.

It seems to be obvious that the overall chance of this comparison being between two genuine messages is:

```
                15        45
```

```
     (1 - ---)*(1 - ---)
          100       100
```

and indeed a table can be drawn up tabulating (in hdB) the overall likelihood of any pair of messages being genuine:

	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%
0%	0	0	-1	-1	-2	-2	-3	-4	-4	-5	-6	-7	-8	-9	-10
5%	0	-1	-1	-2	-2	-3	-4	-4	-5	-6	-6	-7	-8	-10	-11
10%	-1	-1	-2	-2	-3	-3	-4	-5	-5	-6	-7	-8	-9	-10	-11
15%	-1	-2	-2	-3	-3	-4	-5	-5	-6	[-7]	-7	-8	-9	-11	-12
20%	-2	-2	-3	-3	-4	-4	-5	-6	-6	-7	-8	-9	-10	-11	-12
25%	-2	-3	-3	-4	-4	-5	-6	-6	-7	-8	-9	-9	-10	-12	-13
30%	-3	-4	-4	-5	-5	-6	-6	-7	-8	-8	-9	-10	-11	-12	-14
35%	-4	-4	-5	-5	-6	-6	-7	-7	-8	-9	-10	-11	-12	-13	-14
40%	-4	-5	-5	-6	-6	-7	-8	-8	-9	-10	-10	-11	-12	-14	-15
45%	-5	-6	-6	-7	-7	-8	-8	-9	-10	-10	-11	-12	-13	-14	-16
50%	-6	-6	-7	-7	-8	-9	-9	-10	-10	-11	-12	-13	-14	-15	-16
55%	-7	-7	-8	-8	-9	-9	-10	-11	-11	-12	-13	-14	-15	-16	-17
60%	-8	-8	-9	-9	-10	-10	-11	-12	-12	-13	-14	-15	-16	-17	-18
65%	-9	-10	-10	-11	-11	-12	-12	-13	-14	-14	-15	-16	-17	-18	-20
70%	-10	-11	-11	-12	-12	-13	-14	-14	-15	-16	-16	-17	-18	-20	-21

Alexander's description says that the Big Room girl doing the scoring would look opposite 15 and under 45 and get the overall loss which would be -4.

We get -7. This is a massive error and needs investigation. There would seem to be two obvious possibilities:

- Alexander (or more likely his typist) made (another) mistake. The easiest one to swallow is that the CQT message is actually 25% likely to be dummy.
- Maybe losses were not handled in hdB, but another logarithmic system. However, there is a fair bit of other evidence (like the fact that the examples in section 5 came out close to the claimed figures) to suggest that hdB is indeed correct.

Unless proven wrong later, let's assume that Alexander just made a19nother simple typo. (It won't be the last time we have to query the figures in [ALEX46] para. 26.)

Next we need to find the loss due to 'distance' (2.3 in this case). A reconstructed middle-wheel distance chart was shown in section 6.1 above, and we look opposite 2 and under 3 to get a loss due to distance of -2. This agrees with Alexander's description (more evidence that losses are in hdB).

Now for the tetra. The two messages are stated to be in categories XI and XIV, and the tetras appear "in the middle" somewhere (we don't have quite enough evidence to say where, but obviously neither tetra starts within 30 characters of the start or the end of either message). We can't rebuild BP's tetra-chart (the best we could do was shown in section 4.2.2 above, using this example as its source). So we'll have to take it as stated that the tetra is worth 19 + 23, i.e 42hdB total.

Now for the rest of the repeat, with the tetra removed. This comes out as "**15^{xx}/266**" according to Alexander, but here we have a double conflict to resolve:

- Alexander seems to have forgotten to decrement 4 characters for the missing tetra from the overall overlap length. It would seem that "**15^{xx}/262**" is more correct.
- The uncorrected conflict between lines 5 and 6 of [ALEX1946] page 106 leaves us unsure whether in fact the repeat to be evaluated shouldn't be "**14^{xx}/262**" anyway.

As it happens "**14^{xx}/262**" looks more likely in this case. This repeat needs to be looked up on a chart compiled for a loss of -6hdB (losses due to distance plus dummyismus):

Number	Overlap Length											
of repeats	260	261	262	263	264	265	266	267	268	269	270	271
10	-3	-3	-3	-3	-3	-3	-3	-3	-3	-4	-4	-4
11	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	-2
12	1	1	1	1	1	1	0	0	0	0	0	0
13	3	3	3	3	3	3	3	3	2	2	2	2
14	5	5	5	5	5	5	5	5	5	4	4	4
15	7	7	7	7	7	7	7	7	7	7	7	6
16	10	9	[9]	9	9	9	9	9	9	9	9	9
17	12	12	12	11	11	11	11	11	11	11	11	11
18	14	14	14	14	13	13	13	13	13	13	13	13
19	16	16	16	16	16	16	15	15	15	15	15	15
20	18	18	18	18	18	18	18	17	17	17	17	17
21	20	20	20	20	20	20	20	20	20	19	19	19
22	22	22	22	22	22	22	22	22	22	22	21	21
23	25	24	24	24	24	24	24	24	24	24	24	24
24	27	27	27	26	26	26	26	26	26	26	26	26

We get a bonus of 1 for each bigram and thus look up 16/262 and find a score of +9 (Alexander claims +9 too).

We now add up the score for the tetra and the score for the rest of the repeat. Alexander claims:

```
Total score = 42 - 4 - 2 + 9 => 45
```

Because of the use of a scoring chart pre-compensated for a loss of -6hdB, the +9 score for the bulk of the repeat is already correct. However the +42 for the tetra needs to be corrected, and as can be seen above, Alexander merely adds the total loss (-6) to the +42 obtained from the tetra chart.

This is an approximation, albeit a fairly good one. We saw in section 4.4 how to apply a loss to a Bayes Factor. In this case, where the Bayes Factor in question is the overall score for a tetra:

$$B.F = \frac{LT + (1 - L)R}{R}$$

... where L is the probability of the tetra being right, T is the probability of seeing a tetra match between German navy messages and R is the probability of seeing a tetra between random characters. We don't have T and R separated, we have the value of a Bayes Factor for no loss, i.e T/R. Better do some rewriting then:

```
      LT
B.F = --  + (1 - L)
       R

or...

B.F(corrected) = B.F(uncorrected)*L - L + 1
```

We want to work with scores and losses in hdB however, so:

```
Score(corrected) = 20*log(alog(Score(uncorrected)/20)*L - L + 1)

where

L = alog(Loss/20)
```

As long as the absolute loss doesn't exceed about half the score for the tetra, Alexander's approximate result will be right within ±1. For instance, if you apply a loss of -6 to a score of +42, direct addition gives a result of +36. The true answer is +36.065 which is obviously plenty good enough.

(However, if you apply a loss of -42 to a score of +42, direct addition gives a result of 0. It should be +6.)

6.4: Correlating Middle-Wheel and End-Wheel comparisons:

Converting scores to odds for the middle-wheel comparisons involves adding the Bayesian Prior, which in this case is -62. The overall odds for the tetra repeat shown above which scored +45 overall would be alog((45-62)/20) which comes out as 7:1 against.

In addition to noting this score for later use when identifying the alphabet of the middle wheel, the knowledge that it contains concerning the end wheel needs to be transferred to the deciban sheets where it may augment any other information about the end wheel gleaned from end-wheel comparisons.

The closing lines of [ALEX1946] para 26 reveal that in order to do this, BP would correct the raw score by table-lookup and enter the corrected score on the deciban sheets. It would seem that they would apply a loss of (34 - 62)hdB to the raw middle-wheel score before recording it on the deciban sheets. (This loss represents the difference in Bayesian priors between end-wheel and middle-wheel comparisons.) So the conversion chart mentioned by Alexander would tabulate scores with a loss of -28hdB applied. As was argued above in section 6.3, just subtracting 28 would likely not produce accurate results. A portion of a conversion chart for -28hdB loss is reproduced below:

```
 Raw score: | 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49
50 51 52
```

```
------------+-------------------------------------------------------------------
--------
Converts to:|  5  6  6  7  8  8  9 10 10 11 12 13 13 14 15 16 17[18]18 19 20 21
22 23 24
```

The example in [ALEX1946] para 26 notes that a raw score of +45 gets converted by the chart to +18 which is exactly what we see above.

7.0: Scritchmus.

Scritchmus is the art of fitting together all the clues provided by the repeats in order to construct the 'alphabet' for the Enigma machine at grundstellung+1 and grundstellung+2 (in other words for the point where the second and third letters of the indicators had been enciphered).

We can do this because it is known that the three-letter rotor-setting steps forward in a fairly methodical fashion as each letter of the message (or of the indicator itself) was being enciphered. We also know two other properties of the Enigma machine which will be exploited:

- No letter can encipher as itself.
- If 'A' enciphers as 'B' then 'B' enciphers as 'A'.

Of all the component activities of Banburismus as a whole, Scritchmus has been best documented and most easily understood. Both [ALEX1946] and [MAHON1947] carry good examples, and [SALE2000b] also documents the process based on the original example in [ALEX1946].

Scritchmus proceeds in an orderly manner. The deciban sheets are searched for all distances whose scores represent odds of better than 1:1 (i.e with scores >= +34). An attempt is then made to construct the 'end wheel alphabet' by forming 'chains' of end-wheel letters out of these repeats. For instance:

KES = KET+0.8 (10:1 on)
BRT = BQL+2.4 (15:1 on)
AUP = ABT-3.3 (5:1 on)

..would give rise to a chain looking like this:

```
LP..T.......S
    ---------
```

The 8 character distance from 'T' to 'S' is underlined to remind us that it was found from an end-wheel comparison, and that obviously no middle-wheel turnover can have occurred in that stretch because the left and middle wheel indicator letters ('K' and 'E' respectively) haven't changed. This becomes important later...

The chains are then 'scritched' along an alphabet representing the possible plaintext letters. Due to the nice neat stepping of the rotors, this is easily written down as just the letters A-Z in alphabetical order!

We discount all cases where one of the letters in the chain either tries to encipher to itself, or that the reciprocal property of encipherment fails:

```
LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

 LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

  LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

   LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (P <-> E and P <-> S)

    LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

     LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

      LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (S <-> S)

       LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (L <-> H and L <-> T)

        LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

         LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

          LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (L <-> K and L <-> P)

           LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (L <-> L)

            LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

             LP..T.......S
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

S             LP..T.......
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (T <-> S and S <-> A)

.S             LP..T......
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (T <-> T)

..S             LP..T.....
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

...S             LP..T....
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (S <-> D and P <-> S)

....S             LP..T...
```

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (S <-> E and S <-> L)

.....S             LP..T..
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (T <-> L and T <-> X)

......S             LP..T.
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

.......S             LP..T
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

T.......S             LP..
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

.T.......S             LP.
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

..T.......S             LP
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - possible

P..T.......S             L
ABCDEFGHIJKLMNOPQRSTUVWXYZ  - impossible (L <-> S and L <-> Z)
```

So 11 possible plaintext alphabets have just been eliminated just because we had scritched a chain of a mere 4 letters! Obviously, we are still left with 15 possibilities, but it should be possible to eliminate most if not all-but-one of these others by finding other letter-chains and scritching them too.

Because the "Army Wheels" (numbers I thru V) of the supplied set of rotors caused turnovers at different points in their rotations, it is possible to use knowledge of the alphabets to eliminate certain rotors from being at the end-wheel position. For instance, if we knew from other information that the following alphabet was correct:

```
..T.......S             LP
ABCDEFGHIJKLMNOPQRSTUVWXYZ
  ---------
```

...then because the repeat that gave us the 'T' to 'S' distance was an end-wheel comparison, we would know the turnover didn't happen in the gap between them. This would therefore eliminate wheels II and IV from consideration as the end-wheel. (Wheel II turns over between E and F. Wheel IV turns over between J and K.)

BP used the mnemonic "Royal Flags Wave Kings Above" to remember which wheel turned over where. The "Navy Wheels" (VI, VII and VIII) had two turnover notches, one at 'Z' to 'A' (just like wheel V) and another at 'M' to 'N'. Conveniently, the word "Navy" is a mnemonic for this second turnover point.

8.0: A worked example on English messages.

The best way to demonstrate Banburismus is to use it to break a set of messages. Some intercepts do survive from the original German naval Enigma traffic of WWII, but it is easy to create our own with modern computers and Enigma machine simulations. Doing this of course lets us choose the characteristics of our home-grown "intercepts" to suit the job in hand.

A set of 199 suitably enciphered messages are available for download in a choice of two formats:

Enigma "intercepts" for Banburismus attack	
Intercepts (pkzipped for DOS/Windows)	Intercepts (TAR-ed and gzipped for Unix/Linux)

Some aspects of WWII naval Enigma have been preserved, others have not:

- The set of rotors used in the key will contain at least one "navy wheel". This is authentic.
- Adjacent letters of the alphabet will not be steckered together. This is authentic.
- Ten pairs of letters will be steckered, six left unsteckered. This is authentic.
- All messages are 'genuine' (i.e no dummyismus will be required).
- Plaintext of messages is plain English, not abbreviated German.
- The German navy's 250 character limit for individual messages is not enforced (this increases the chance of finding useful mid-wheel comparisons).

8.1: Charts for the English language - ROMSing.

From all the theory work above (sections 3 and 4) re-deriving German charts, it is obvious what is needed in order to generate charts suitable for scoring English language messages.

We need the probabilities of monogram, bigram (etc) repeats between typical messages. This has been done for some messages typical of the ones used to create the downloadable cryptotext above. Obviously, the hard work was done with a modern computer, but it seems from comments by Alexander and Good that the statisticians of BP used the Hollerith card-processing machinery of the era to achieve the same goal. They called the process "ROMS-ing" where "ROMS" stood for "Resources of Modern Science".

(Maybe an enthusiast for Hollerith machinery can shed some light on how they might have done this - I know nothing about it.)

The results of ROMSing were:

- Monogram repeats are found to be likely every 18 characters, not 17 as was the case with the WWII messages. The score for a monogram repeat is therefore 20*log(26/17) which is +3.7
- Bigram repeats score +10
- Trigram repeats score +25
- Tetras score +38
- Pentagram repeats score +54
- Hexagram repeats score +73
- Heptagram repeats score +95
- Octagram repeats score +116
- Enneagram repeats score +138

A set of 'postscript' charts are available for download in a choice of two formats:

Scoring Charts for Banburismus attack	
Charts (pkzipped for DOS/Windows)	Charts (TAR-ed and gzipped for Unix/Linux)

The filenames for the charts are in the form "chart-s-l.ps" where 's' is the starting overlap length, and 'l' is the loss. They should print out correctly regardless of whether you have European A4 paper or North American style 'letter' paper.

8.2: Compiling the scores.

The charts are used in pretty much the same way as were BP's original WWII charts (see sections 4, 5 and 6 for details). The primary difference is in the "bonus" rows to be awarded for bigrams and trigrams. The different characteristics of the modern messages call for a bonus of +2 rows for bigrams and +6 for trigrams.

After many happy hours slaving away in the garden shed with your home-made Banbury sheets, light table and salvaged Hollerith machinery :-) you should have a score sheet something like this and a set of deciban sheets something like this . We'll use the score sheet later when attacking the middle wheel, but for the end-wheel we want to take the deciban sheet and extract just those entries with scores higher than +30. This will give us a "good fit list" for the end-wheel, where the odds of each distance being right are about evens or better:

------- Deciban Sheet Summary -------

Score	Distance	Odds
[150]	G=K+4	Certain!
[141]	B=N-24	Certain!
[80]	M=Q+16	200:1 on
[61]	C=Q+18	22:1 on
[53]	J=L+24	9:1 on
[51]	C=M+2	7:1 on
[47]	G=V+9	9:2 on
[46]	A=Z-10	4:1 on
[43]	N=R-23	3:1 on
[43]	G=O+23	3:1 on
[41]	D=H+10	9:4 on
[40]	C=J-19	2:1 on
[39]	Q=T+4	9:5 on
[39]	M=N+10	9:5 on
[39]	L=P+7	9:5 on
[35]	E=N-8	Evens
[34]	R=Y+17	Evens
[31]	G=T-4	4:3 against

[31]	E=O-12	4:3 against
[30]	W=X+14	3:2 against

8.3: Letter Chains, Scritchmus and the End-Wheel Alphabet.

Starting with the entries that get the most certain odds, we can compose a number of end-wheel letter-chains as follows (sorted by length and within that, the overall most likely-to-be-correct):

m.c...t...q	The "M-C" distance is obtained from C=M+2 at 7:1 on, with both the C=Q+18 and M=Q+16 agreeing as to the placement of the letter 'Q' (total score for the triad being 51+80+61 = 192, i.e appx 4.0e09:1 on!). The "T-Q" distance is a bit weak in comparison at only 9:5 on, but we can probably get away with it...
v....k...g..o	The "K-G" distance is obtained from G=K+4 (certain), with the letter 'V' placed as a result of G=V+9 at 9:2 on. Letter 'O' can be placed as a result of G=O+23 at 3:1 on.
r..n.b	The "N-B" distance is obtained from B=N-24 (certain), with the letter 'R' placed as a result of N=R-23 at 3:1 on. Nothing much else fits safely into this chain.
j.l	This distance is obtained from J=L+24 at 9:1 on.
a.........z	This distance is obtained from A=Z-10 at 4:1 on.
h.........d	This distance is obtained from D=H+10 at 9:4 on.

We notice that unlike the examples of scritchmus shown in [ALEX46] and [MAHON47], the best letter-chains we can assemble are not nice and long with lots of letters present! We have instead got several short chains which will make the scritchmus procedure long and difficult. This is however probably exactly the sort of situation in which Hut 8 found themselves on most days.

Starting from the top of the list, we find the following possible plaintext alphabets for the first chain:

```
.....m.c...t...q.........
abcdefghijklmnopqrstuvwxyz
cdefghijklmnopqrstuvwxyzab
defghijklmnopqrstuvwxyzabc
```

```
ghijklmnopqrstuvwxyzabcdef
klmnopqrstuvwxyzabcdefghij
pqrstuvwxyzabcdefghijklmno
qrstuvwxyzabcdefghijklmnop
stuvwxyzabcdefghijklmnopqr
tuvwxyzabcdefghijklmnopqrs
uvwxyzabcdefghijklmnopqrst
wxyzabcdefghijklmnopqrstuv
yzabcdefghijklmnopqrstuvwx
zabcdefghijklmnopqrstuvwxy
```

Notice that we're already down to only 13 possible plaintext alphabets from 26, and we've only got 4 letters placed. However, for each of these possible alphabets we have to "scritch in" the second chain.

The first valid position for the second chain is as follows:

```
.v...mkc..gt.o.q..........
abcdefghijklmnopqrstuvwxyz
yzabcdefghijklmnopqrstuvwx
```

We're down to only 2 possible plaintexts with the second chain positioned like this, but unfortunately, the second chain can also be placed thus:

```
....vm.c.k.t.g.qo.........
stuvwxyzabcdefghijklmnopqr
```

For each valid placing of the second chain (there are 12 of them), a small set of possible plaintext alphabets is valid. In fact there seems to be a total of 23 such alphabets.

For each of these 12 valid placings of the second chain, we have to scritch the third chain in. And at this point we hit a snag. There is no valid way to do this. Evidently, one of the distances used to compile the first three chains is a dud. It's difficult to see how to proceed any better than to remove the distance with the worst odds (G=O+23 at 3:1 on) and try scritching again.

The second chain now looks like this:

v....k...g	The "K-G" distance is obtained from G=K+4 (certain), with the letter 'V' placed as a result of G=V+9 at 9:2 on.

To cut a long story short, we find that doing this frees the deadlock, and by the time we've got all 6 chains scritched in, we'll find that there are only four possible alphabets left:

```
    j.lakm.cg.ht.zrq.n.bd....v
    cdefghijklmnopqrstuvwxyzab

    j.l.kmacg.ht..rqzn.bd....v
    cdefghijklmnopqrstuvwxyzab

    j.l.km.cg.hta.rq.n.bd.z..v
    cdefghijklmnopqrstuvwxyzab
```

```
j.l.kmzcg.ht..rq.n.bd.a..v
cdefghijklmnopqrstuvwxyzab
```

Only one of these can be right. It helps to "fill in" these alphabets by means of reciprocal letter-pairs where possible though:

```
jwlakm.cgeht.zrq.n.bd..pfv
cdefghijklmnopqrstuvwxyzab

jwl.kmacgeht..rqzn.bd..siv
cdefghijklmnopqrstuvwxyzab

jwl.km.cgehta.rq.n.bd.zyov
cdefghijklmnopqrstuvwxyzab

jwl.kmzcgeht..rq.n.bd.aiyv
cdefghijklmnopqrstuvwxyzab
```

We notice that:

- All of the alphabets contradict the weak distance T=G+4 (3:2 against) which isn't very important. They also all contradict the stronger distance M=N+10 (9:5 on).
- All of the alphabets "pick up" the distance C=J-19 (2:1 on) and the weaker distance E=N-8 (evens). This is good for morale, but doesn't tell us anything useful!
- Only the third alphabet "picks up" the weak distance R=Y+17 (evens).
- The top alphabet contradicts both the distances L=P+7 (9:5 on) and the weak distance W=X+14 (3:2 against). The other three all permit L=P+7, making the letters X <-> P reciprocal. This then "picks up" the weak distance W=X+14 (3:2 against), but contradicts the similarly weak distance E=O-12 (4:3 against).

It is fairly obvious that the top alphabet of the four picks up the least number of lesser distances listed and can probably be discounted. It is however very difficult to decide between the other three at this point, though alphabet number three of the four is slightly in the lead due to picking up R=Y+17.

We really have little alternative now than to work through the deciban sheets totting up the scores in favour of and against each of alphabets two, three and four above. If we keep relative score rather than absolute, then we'll only be interested in totting up scores where they count for some alphabets and against others. So far, by that assumption, alphabet three starts with +34 and the other two start at -34 for the distance R=Y+17 (evens):

	Alphabets		
	2	3	4
So far:	-34	+34	-34
K=S+1:	-29	+29	+29
Q=U-17:	-23	+23	-23
N=Z+11:	-23	-23	+23
C=Z-15:	-23	+23	-23

O=Z+2: -18 +18 -18
O=Y-25: -18 +18 -18
A=P-9: -17 +17 -17

It would seem at this point to be pretty obvious that the third of our four alphabets has easily won the race with no real chance of either of the other contenders making up lost ground. We can fill in all the missing letters of this alphabet as follows:

```
    jwlskmucgehtaxrqfnibdpzyov
    cdefghijklmnopqrstuvwxyzab
```

Now to work out what wheel(s) of the Enigma machine can be eliminated from the end-wheel position.

8.4: Royal Flags Wave Kings Above (not forgetting the Navy).

Now we want to plot onto this alphabet all the distances known from the end-wheel comparisons. It is important to use only the actual end-wheel comparisons listed on the deciban sheets - many deciban sheet entries derive from mid-wheel comparisons and are useless in this context.

Because the middle letters of their indicators didn't change we know that no middle-wheel turnover happened in those distances found by true end-wheel comparisons:

```
    2     4   N    1     3   5,N
 jwl|skmuc|geh|taxr|qfnib|dpzy|ov
 cde|fghij|klm|nopq|rstuv|wxyz|ab
            +---------+               (from BJE=BJN-0.8)
<-+              +--------------->    (from MLX=MLW-0.14)
       +-+                            (from NFM=NFC-0.2)
               +----+                 (from NLT=NLQ-0.4)
      +----+                          (from UOK=UOG-0.4)
```

The first two of these distances have killed off any chance that we've got a "Navy Wheel" in the end position. E=N-8 denies a turnover between M/N, and X=W-14 denies a turnover beween Z/A, both of which need to be possible for a Navy Wheel. Between them (with help from T=Q-4) they also deny wheel I with its turnover between Q/R.

X=W-14 in denying the Z/A turnover also removes the possibility of wheel V being the end-wheel. It also denies the V/W turnover of wheel III.

Wheel IV with its turnover between J/K is denied by the K=G-4 distance.

Only wheel II with the turnover between E/F is possible.

Had we been struggling to find enough end-wheel comparisons with which to isolate the identity of the end-wheel, we could have pulled in two more bits of evidence in the shape of the mid-wheel comparisons YJR=YKN+0.23 and SSY=SCR-0.17. (The unique feature of them is that the distances indicated are less than 26 characters.)

The first of these extra clues is saying that there definately is a middle-wheel turnover in the 23 character stretch R=N+23. Since we know that we don't have a Navy Wheel as the end-wheel, we can safely say that there isn't a turnover in the 3 character stretch R=N-3! Likewise, we can say that there isn't a turnover in the 9 character stretch Y=R+9.

```
    2    4   N    1     3   5,N
 jwl|skmuc|geh|taxr|qfnib|dpzy|ov
 cde|fghij|klm|nopq|rstuv|wxyz|ab
                  +---+                (from R=N-3)
                  +----------+         (from Y=R+9)
```

Actually, in this case these extra distances don't add anything that we didn't already know. Wheels I and III are now denied a bit more thoroughly than before, but that is all. There will be cases however where the extra information gleaned could be critical.

8.5: The Middle-Wheel Alphabet.

First thing to do is to search through the score sheet looking for middle-wheel comparisons whose scores are better than about 30 and whose end-wheel letter distances actually agree with the end-wheel alphabet that we found above.

We will therefore discard known dud middle-wheel comparisons like the infamous UPO=UOG-1.23 above. It might have a score of +71, but the O=G-23 distance is obviously wrong.

Normally, the Bayesian Prior for mid-wheel comparisons is -62, but if we select just those comparisons which are already known to have correct end-wheel distances, then as Alexander says "it vastly increases the chance of the tetra being right". In fact, it would seem that we can consider using a Bayesian Prior of -34 for these selected tetras.

Additionally, we have to do a bit of work on each comparison to determine the exact motion of the middle wheel and thus the middle-wheel distance. The reason for this is that the middle wheel steps only because the end-wheel turnover notch makes it move. If we don't know the location of the end-wheel turnover notch we can't attack the middle-wheel alphabet.

Take for example the BIN=BLB+2.24 tetra. We know the end-wheel alphabet looks like this:

```
   T/O
 jwl|skmucgehtaxrqfnibdpzyov
 cde|fghijklmnopqrstuvwxyzab
                   ^ ^
                  e s
                  n t
                  d a
                    r
                    t
```

To get from 'B' on this alphabet to 'N' in +2.24 letters, we must pass the turnover point three times. So from the mid-wheel alphabet's point of view, I=L+3.

Taking all this into account, our new list looks like this:

Score	Repeat		Odds	Mid-wheel distance
[169]	BIN=BLB+2.24	$37^{83XXX}/382$	Certain!	I=L+3
[108]	FYQ=FZM-1.16	$17^{73}/253$	Certain!	Y=Z-2
[89]	NHQ=NFC-1.18	$33^{4XX}/290$	560:1 on	H=F-2
[81]	GOL=GHJ-6.24	$22^{53XXX}/208$	224:1 on	O=H-7
[75]	UOG=UQV+1.9	$29^{4XXX}/366$	112:1 on	O=Q+2
[74]	EUA=EAZ-4.10	$24^{6}/353$	100:1 on	U=A-4
[71]	YJR=YKN+0.23	$20^{43X}/256$	71:1 on	J=K+1
[69]	WGH=WVD-5.10	$27^{43X}/303$	56:1 on	G=V-5
[68]	RCC=RYJ-6.19	$15^{6X}/227$	50:1 on	C=Y-6
[67]	YSP=YEL-3.7	$29^{4XX}/321$	45:1 on	S=E-3
[62]	SSY=SCR-0.17	$18^{4XXX}/253$	25:1 on	S=C-1
[51]	NHQ=NDU-2.17	$22^{4}/265$	7:1 on	H=D-3
[46]	UPO=UBY-4.25	$26^{4X}/385$	4:1 on	P=B-5
[44]	KKA=KJP-1.9	$20^{4}/296$	3:1 on	K=J-1
[37]	OUP=OCR-6.19	$20^{4}/277$	4:3 on	U=C-7

8.6: Letter Chains, Scritchmus and the Middle-Wheel Alphabet.

Starting with the entries that get the most certain odds, we can compose a number of middle-wheel letter-chains as follows (sorted by length and within that, the overall most likely-to-be-correct):

`q.o......h.fd`	The "O-H" distance is obtained from O=H-7 at 224:1 on, with the H=F-2 (560:1 on) and H=D-3 (7:1 on) and O=Q+2 (112:1 on) distances providing the rest of the picture.
`sc.e...y.z`	The "S-C" distance is obtained from S=C-1 at 25:1 on, with the letter 'E' placed as a result of S=E-3 at 45:1 on. "C=Y-6" at 50:1 on fits in nicely here too. "Y=Z-2" (certain) can therefore be pulled in too. "U=C-7" at only 4:3 on is a bit feeble, so we won't add it here. We'll use it as evidence in favour of alphabets later (or against them of course).
`l..i`	This is obtained from I=L+3 (certain). Nothing else fits though.
`kj`	The "K=J-1" distance is obtained from J=K+1 at 71:1 on backed up by K=J-1 at 3:1 on.

	Total score = 44 + 71 = 115, i.e over 110000:1 on.
`u...a`	This distance is obtained from U=A-4 at 100:1 on.
`g....v`	This distance is obtained from G=V-5 at 56:1 on.

As with the examples of scritchmus shown above, we have to deal with quite a few short sparse chains rather than nice, long, well-populated ones! However, even after only scritching the first two chains, we'll find ourselves in the situation where the chains only fit together in two possible ways yielding two possible alphabets:

```
q.o....schefd.y.z.........
fghijklmnopqrstuvwxyzabcde

qeo..y.z.h.fd...........sc
klmnopqrstuvwxyzabcdefghij
```

Of course, as before with the end-wheel alphabet, once you start to scritch the two-letter chains in, things tend to explode a bit before reducing back to sensible small numbers of possibilities.

Amazingly though, the "L-I" chain expands the above to yield just four possible alphabets, adding in "K-J" expands the list to six, adding in "U-A" reduces that to 2, and "G-V" further reduces it to.... just one!

```
quol.aischefd.ygz...vkj...
fghijklmnopqrstuvwxyzabcde

and with the known blanks filled in by reciprocity:

quolbaischefdmygz..tvkjnrp
fghijklmnopqrstuvwxyzabcde
```

We notice immediately that this alphabet picks up U=C-7 for a score of +37 and also picks up P=B-5 for a score of +46.

8.7: Royal Flags Wave Kings Above (not forgetting the Navy). Again.

Just as with the end-wheel, we want to plot onto this alphabet all the distances known from the mid-wheel comparisons. All mid-wheel comparisons are valid for this as in all cases they derive from cases where the slow wheel didn't experience a carry:

```
 2     4   N    1     3   5,N
 |quolb|ais|chef|dmygz|xwtv|kjnrp
 |fghij|klm|nopq|rstuv|wxyz|abcde
     +----+                           (from I=L+3: refutes wheel IV)
                   +-+                (from Y=Z-2)
             +-+                      (from H=F-2)
    +--------+                        (from O=H-7: refutes wheel IV and all Navy
wheels)
  +-+                                 (from O=Q+2)
```

```
   +-----+                            (from U=A-4: refutes wheel IV again!)
                            ++        (from J=K+1 and K=J-1)
                    +------+          (from G=V-5: refutes wheel III)
            +-------+                 (from C=Y-6: refutes wheel I)
          +----+                      (from S=E-3: refutes Navy wheels again.)
          +-+                         (from S=C-1: refutes Navy wheels again
again.)
             +----+                   (from H=D-3: refutes wheel I again.)
<-----+                         +->   (from P=B-5: refutes wheel II)
```

The last of these distances refutes wheel II, though we knew that anyway because that's established as the end-wheel from earlier work.

Only wheel V with the turnover between Z/A is possible.

Because we're assuming in this document that the German navy insisted that at least one "Navy Wheel" was always to be present in the Walzenlage, we are now in the rare and lucky position of being able to say that only wheel orders 652, 752 and 852 need to be tried in the bombes! At 15 minutes per run, that means we should now be only 45 minutes away from finding the key for the day.

9.0: The Bombe.

The 'Bombe' was a special-purpose electromechanical key-solving machine. It was programmed by plugging cables into sockets on the rear and setting the relative positions of many drums representing the wheels of Enigma machines on the front. It had no "memory" in the normal sense of the word, yet was for its intended purpose a very powerful parallel processing machine.

Hobbyists at Bletchley Park museum are currently rebuilding a bombe . The original British bombes were mostly destroyed after WWII was over, though some were allegedly moved to the new codebreaking centre "GCHQ" in Cheltenham and must have been used against post-WWII rotor machines.

Apparently one USN bombe survives in the U.S, but it isn't working.

The British "Bombe" was quite radically different from the earlier Polish "Bomba" though it would seem that the name and some fundamental ideas of the Polish machine led directly to the British effort. Credit must be given to the Poles for their outstanding efforts all through the 1930s in developing attacks on Enigma machines. Their "Bomby" solved a different sort of problem (repeating letters in old-style indicator groups), but had many ideas in common with the more general British machine.

Plenty has been written on how the Bombes worked, and how they were set up. Check out the American report on the British Bombe for instance. There's far too much information to be reproduced here!

9.1: Bombe menus.

Normally, the programming (known as a "menu") for the Bombe would be produced from a crib - an educated guess as to part of the contents of one of the messages. However, in the early days of Banburismus there were few or no cribs available.

If (as is true in the case above) a fairly complete alphabet of the machine is known for the middle and end wheels, then a menu can be constructed from that knowledge alone. The bombe ends up with all its scramblers set to positions 2 and 3, and produces (if we're lucky) a drop representing the basic key for the day.

However, a bombe can't solve for the Ringstellung, and this then means that we still can't yet use that key directly to read the messages.

As usual, examples are better than words. The mid- and end-wheel alphabets for the machine look like this:

```
abcdefghijklmnopqrstuvwxyz
--------------------------
kjnrpquolbaischefdmygzxwtv    <-  middle wheel (at grundstellung+1)
ovjwlskmucgehtaxrqfnibdpzy    <-   end wheel (at grundstellung+2)
```

We can construct a menu as follows. It just so happens that the letters arrange themselves into two rings, one needing 18 scramblers to implement, the other requiring 8:

```
    2   3   2   3   2   3   2   3   2   3   2   3    2
+-A---K---G---U---I---L---E---P---X---W---D---R---Q---+
|                                                     |
|                  3                2   3   2   3     |
+---------------------------------O---H---M---S---F---+

     2   3   2   3    2
+--Z---V---B---J---C---+
|                      |
|  3         2   3     |
+----------Y---T---N---+
```

This is using an ASCII-art representation of the menu nomenclature described by Derek Taunt in [HINS1993]. Taking the letters of the second ring for example, the menu says that letter 'Z' enciphers as 'V' on an Enigma machine set at Grundstellung+2, then that letter 'V' would encipher as 'B' on a machine set to Grunstellung+3. Letter 'B' would encipher as 'J' back at Grundstellung+2 etc etc. Eventually the loop closes after 8 of these transformations.

Assuming that there isn't a turnover between Grundstellung+2 and +3, then the bombe ought to find all possible settings of the Enigma machine where these letter relationships can happen. Hopefully, there will only be one such solution amongst wheel orders 652, 752 and 852.

We apply power to letter 'Z' of cable 'A' (which therefore also energises letter 'A' of cable 'Z' via the diagonal board). The test register is connected to cable 'N'. There are many more possible ways to power up this menu, but this is what the author chose to do....

We get exactly one drop:

Wheel Order: 6,5,2
Ringstellung: C,A,N
Steckers: AF BJ CI ER HK LZ MQ NT PX UY
Self-Steckers: D,G,O,S,V,W
Grundstellung: N,A,F

Hut 8 would consider this a perfect solution as normal Naval procedure would dictate the 10 stecker-pairs and 6 self-steckers seen here. Not only that, but it's the only solution found by the bombe!

The problem with this solution is that a bombe cannot solve for a unique Ringstellung/Grundstellung combination. A Ringstellung of "C,B,N" and Grundstellung of "N,B,F" would be valid, as would "C,B,P"/"N,B,H". There are in fact about 17000 possible Ringstellung/Grundstellung combinations that will work! The only ones that can be ruled out are the few that would have caused a turnover between Grundstellung+2 and +3. With wheel-order of 6,5,2 the use of Grundstellung "N,A,F" means that we get the longest ciphering run possible before a turnover, but that's just the canonical solution offered by the author's bombe simulator.

9.2: Getting into the messages themselves.

When you have a crib for a message, you can determine the correct Ringstellung by working forwards in the message until it breaks into gibberish and that's the position on the real key where the turnover happened. The end-wheel Ringstellung/Grundstellung combination for the bombe's solution of the key can then be twiddled to match the real key, and at some point the turnover of the middle wheel will make itself known and that too can be corrected.

We get no such help in this situation. Our only way to get into the messages themselves will be to compile an "eins catalogue".

9.3: The "Eins Catalogue".

When you know a key as completely as we do here, we can use a machine that Hut 8 called "The Baby" (mentioned by Joan Murray in [HINS1993]) to generate the encipherment of any likely four-letter group at every one of the 17576 possible message-settings of the Enigma machine. When attacking naval Enigma, the standard technique was to use the word "eins" for this task - hence the name "Eins Catalogue". The "Baby" would encipher "eins" and punch the results on Hollerith Cards, these would then be sorted into order and a given message would be searched laboriously for any occurrences of any of the encipherments listed in the catalogue.

"Eins" was a good group to use against naval Enigma because numbers had to be spelt out in full, and the digit '1' occurs more often than any other digit in naturally-occurring numbers. (Obscure, but true). Also the letter-group "eins" crops up in many normal German words.

"Eins" would however be a rotten candidate group for attacking messages in English! Unfortunately the spelling of the digit '1' in English is a 3-letter group, and thus not useful for cataloguing. This is because there are only 17576 possible arrangements of three letters, and an Enigma machine will encipher any given 3-letter group to yield almost all of them! So our catalogue lookups will almost always succeed and we would spend a huge amount of time checking each one. A four-letter group found in a message on the other hand will only match a catalogue based on four-letter groups about once in 26 letters which amounts to about 7 or 8 possible 'hits' in a typical 200-character message.

It seems (from testing some candidate text) that for generic English messages the four-letter group "OTHE" is a good candidate. This catches all the cases of words ending in 'O' followed by "the" or "them" or "they" etc etc.

It just so happens however that the senders of the messages we're attacking in this document have made it a bit easy for us by finishing every sentence with "STOP" (like telegrams - remember them?). We shall therefore compile a "Stop Catalogue" and use it to examine one of our longer messages at random:

```
GHBZFQUURKEKJDBUOETKZEJKZNWCUZEZAWMBKMJTDECZXBNUTCHXLOONBUWPHPDYGKODHGWQNLPLAKKX
BVZKLSMRGZQOKDOBPVNQURQDENWUWKTUJJDJLUBZJSQYFYOXUGJSSFITMUTXMTDJATTIFAJSQMVZDTLU
RJEPSNGAIGBUBKZRFTKRHRKTXVHQIKEXEZRMRHGRKSUMQNJQHJMEZWTIGCIQMFUYAOTNYDQZEQASOJCV
QVQSVMSOKRWRRXKCEJTLCIZOPQMSAYNREGUXBYHBQLFTNASQOSCFCTUBDTZUIVHDXNWDKCIYXUKZBRYJ
SNDLHOEUYPCCUMZKIQVDRHXBQPGEGLPBLJUGYEHCPFLBTCQBDWAELGGXWCZCKHPFOSXSAVYKGUYCCVVO
MVAUHCOUXYXRIHKRFFICQAXGLPANBZKLSTFZQWBAYFBKHZMEMYAEGIBJJNDCUC
                                        ^^^^
```

Above is one of our messages, indicator HUC (478 characters long, so we'll have a good chance of witnessing a mid-wheel turnover somewhere). We could proceed by looking up GHBZ in our catalogue, followed by "HBZF", followed by "BZFQ" etc.

In the particular case of a "STOP" catalogue however, we could make things a bit more easy for ourselves by starting at (say) position 40 of the message (with group "DECZ"). This is because we know that the most likely occurances of "STOP" is at the ends of sentences and that sentences won't end too close to the start of the message itself.

The first match past character 40 is "JSQY" for a message-setting of "AGJ" at character number 120 of the message. We'll backstep this message-setting until we get to "AGF" (which is the furthest back we can go before hitting our end-wheel turnover). We then plug the relevant ciphertext into an Enigma machine and get:

```
LUBZJSQYFYOXUGJSSFITMUTXMTDJATTIFAJSQM
TMENSTOPWXCCKUGUWJZZZEPIGNOTBEATTRIKTY
    ^^^^
```

This is quite possibly a hit! (Lucky too - normally we'd expect to have had to check out a few false alarms in the 80 positions we've checked so far).

We can think up all sorts of sentences finishing "t men." (absent men?, diligent men?) and possibly the "W" following the "STOP" might be genuine, but the "X" after that certainly isn't. So the enciphering key did an end-wheel turnover immediately after the "STOP" or possibly immediately after "STOPW".

We might get a clue looking at the decipherment from 26 characters after the "TMENSTOPW", which is giving us "OTBEATTRI" (part of "not be attributed"?).

It's a bit weak, but let's assume that the "W" following "STOP" is genuine, and we want to manipulate our key's ringstellung so that an end-wheel turnover happens on the letter after that. So we want the end-letter of the message- setting to be "E" on the character marked below:

```
LUBZJSQYFYOXUGJSSFITMUTXMTDJATTIFAJSQM
TMENSTOPWXCCKUGUWJZZZEPIGNOTBEATTRIKTY
         ^
```

Remember that an Enigma machine does the rotor movement **before** deciphering/enciphering a letter. So the "E" will change to "F", carrying the middle rotor over as it does so **then** the "Y" will be deciphered to whatever.

This means that the message key for the above fragment needs to be changed to "AGV" and we can do this without disturbing the rotor positions be stepping the ringstellung by as many letters as we want the message-setting to change.

If the message setting for the end-wheel changes from "F" to "V", then the ringstellung for the end-wheel must change from "N" to "D". We try again:

```
LUBZJSQYFYOXUGJSSFITMUTXMTDJATTIFAJSQM
TMENSTOPWHATHEACHIEVEDCANNOTBEATTRIBUT
```

Evidently, that sorts out the end-wheel ringstellung! In order to attack the mid-wheel ringstellung we'll have to backstep our message-setting to try and decode the whole message from the start. "AGV" is the message setting for this fragment starting at character 116. That's "4.12" in BP's "alphabet.letter" nomenclature. By careful counting backwards from "AGV" and taking care of the turnovers for both end- and middle-wheels we find that "ACJ" might decode the whole message (the middle wheel didn't hit its turnover notch in the stretch "G" back to "C"):

```
GHBZFQUURKEKJDBUOETKZEJKZNWCUZEZAWMBKMJTDECZXBNUTCHXLOONBUWPHPDYGKODHGWQNLPLAKKX
BVZKLSMRGZQOKDOBPVNQURQDENWUWKTUJJDJLUBZJSQYFYOXUGJSSFITMUTXMTDJATTIFAJSQMVZDTLU
RJEPSNGAIGBUBKZRFTKRHRKTXVHQIKEXEZRMRHGRKSUMQNJQHJMEZWTIGCIQMFUYAOTNYDQZEQASOJCV
QVQSVMSOKRWRRXKCEJTLCIZOPQMSAYNREGUXBYHBQLFTNASQOSCFCTUBDTZUIVHDXNWDKCIYXUKZBRYJ
SNDLHOEUYPCCUMZKIQVDRHXBQPGEGLPBLJUGYEHCPFLBTCQBDWAELGGXWCZCKHPFOSXSAVYKGUYCCVVO
MVAUHCOUXYXRIHKRFFICQAXGLPANBZKLSTFZQWBAYFBKHZMEMYAEGIBJJNDCUC
NEVERTHELESSHISBARBAROUSCRUELTYANDINHUMANITYWITHINFINITEWICKEDNESSESDONOTPERMITH
IMTOBECELEBRATEDAMONGTHEMOSTEXCELLENTMENSTOPWHATHEACHIEVEDCANNOTBEATTRIBUTEDEITH
ERTOFORTUNEORTOGENIUSSTOPINOURTIMESDURINGTHERULEOFALEXANDERVIOLIVEROTTODAFERMOHA
VINGBEENLEFTANORPHANMANYYEARSBEFOREWASBROUGHTUPBYHISMATERNALUNCLEGIOVANNIFOGLIAN
IANDINTHEEARLYDAYSOFHISYOUTHSENTTOFIGHTUNDERPAOLOVITELLITHATBEINGTRAINEDUNDERHIS
DISCIPLINEHEMIGHTATTAINSOMEHIGHPOSITIONINTHEMILITARYPROFESSION
```

As it happens, this has successfully deciphered the whole message without us encountering the mid-wheel turnover. It lets us know the message-setting according to our version of the key, and we can sort out the middle- and slow-wheel ringstellung from deciphering the indicator (which was "HUC") on the grundstellung discovered by the bombe.

The grundstellung found by the bombe had been "NAF", but this changes to "NAV" to take into account the change we've already made to the end-wheel ringstellung. If we decipher "HUC" on this grundstellung we get "RGJ" yet we thought the message-setting was "ACJ". The third letter is correct (as it must be because we've fixed the ringstellung for that wheel). But the mid wheel on our key indicates "C" when it should be "G". So the our ringstellung is 4 characters behind where it should be (it was "A" but it evidently should be "E"), and the grundstellung needs the same offset applied.

Likewise the slow wheel on our key indicates "A" when it should be "R". So our ringstellung is 17 characters behind where it should be (it was "C" but evidently should be "T").

To sum up, the corrected key for these messages is:

Wheel Order: 6,5,2
Ringstellung: T,E,D
Steckers: AF BJ CI ER HK LZ MQ NT PX UY
Self-Steckers: D,G,O,S,V,W
Grundstellung: E,E,V

For proof, we can directly decode some other random message (say, the one whose indicator was "PDD"):

AFTER THIS LUCCA AND SIENA YIELDED AT ONCE PARTLY THROUGH HATRED AND PARTLY THROUGH FEAR OF THE FLORENTINES AND THE FLORENTINES WOULD HAVE HAD NO REMEDY HAD HE CONTINUED TO PROSPER AS HE WAS PROSPERING THE YEAR THAT ALEXANDER DIED FOR HE HAD ACQUIRED SO MUCH POWER AND REPUTATION THAT HE WOULD HAVE STOOD BY HIMSELF AND NO LONGER HAVE DEPENDED ON THE LUCK AND THE FORCES OF OTHERS BUT SOLELY ON HIS OWN POWER AND ABILITY (from "The Prince" by Niccolo Macciavelli).

9.4: Notes.

The above is possibly the first Banburismus attack performed on a real set of messages with real scoring tables since the decline of Banburismus at BP in 1943. It is possibly the first such attack on English text since Alan Turing himself tested his ideas in 1940 or 1941 (unlikely that he used German messages - he probably got someone to generate a set of messages like these ones, in English for convenience).

The above example is not a fake, and the author did not know the key during the procedure. The key used was generated by computer. It was sheer chance that a navy wheel did not appear as either the middle or end wheels, thus cutting the number of possible wheel orders to three.

By more luck, no turnovers happened in the two rotor positions straight after the grundstellung. Had such a thing happened, it would have complicated things somewhat, but Banburismus would have prevailed. It would however have complicated the description somewhat!

11.0: Dummyismus:

The transmission frequencies, callsigns and message keys - even the morse code "fist" of an operator - were enough for BP to attach a "dummy probability" to each message. The upshot was that all or part of a message would be allocated a "loss" (expressed in hdB) which was due to this chance that it was bogus. [ALEX1946] doesn't explain how BP aquired the knowledge that led to the construction of the relevant "dummy charts" but chances are that they were based initially on experience gained with the material from the 'Narvik' and U110 pinches, and then carefully kept up to date as each new day was broken.

We cannot expect to recreate BP's dummyismus, and even if we did, it would be useless. The system has to be totally tuned to the statistical properties of the incoming messages.

12.0: Source Material:

Most of the historical detail for this file comes from [ALEX1946], a report written by C. H. O'D. Alexander (yes, the famous chess grandmaster) in 1946(?) just as Bletchley Park was being disbanded and reformed as GCHQ.

More historical detail is available in [MAHON1947], though the scoring system is not well documented in that paper. It is useful primarily as it contains an alternative description of scritchmus to the one featured in [ALEX1946].

[HINS1993] is useful mostly for interviews with some of the people who used Banburismus for real, but does include a few technical comments at pages 156 & 157 (of the paperback) from Jack Good one of the people who helped develop Bayesian techniques after WWII and start making it more well-known than it had been previously).

[KAHN1996] is great for a general overview of the Battle of the Atlantic but makes just a few passing references to Banburismus (pages 141 - 143), with no actual details of how it was done. It does however give clues as to the sizes of banbury sheets.

Tony Sale's lecture notes on Naval Enigma [SALE2000a] and [SALE2000b] give various insights into the workings of Hut 8, and discuss Banburismus. They are much more easy to find than [ALEX1946], and are recommended as an introductory read. However, the Banburismus details in [SALE2000a] and [SALE2000b] appear to be gleaned verbatim from [ALEX1946], and no theory is presented.

For the theory stuff, [BERRY1996] is a good, easy-to-read introduction to the essentials of Bayesian Statistics. [LEE1989] goes into the field in rather more depth than we need for this document, but read it if you want to explore Bayesian techniques further.

[WELCH1997] is the paperback revised edition of Welchman's 1982 book of the same name. It deals primarily with Army and Luftwaffe Enigma and of course Welchman was (with Turing) the inventor of the British Bombe. Welchman's reporting of the comments made by the Polish cryptanalysts lead to deductions about Banbury sheets in section 1 of this document.

Ralph Erskine has commented that [CLIFF1943] and [USN1943] may be the same document, or at best two different versions of essentially the same document.

13.0: Bibliography:

ALEX1946: Alexander C. Hugh. O'D.: "Cryptographic history of work on German Naval ENIGMA", Crown Copyright 1946(??), Public Record Office, Kew, Surrey, HW 25/1.

BAUER2007: F. L. Bauer: "Decrypted Secrets: Methods and Maxims of Cryptology", (Berlin: Springer, 2007 (4th ed.)).

BERRY1996: Berry, Donald. A: "Statistics - A Bayesian Perspective", Duxbury Press 1996, ISBN 0-534-23472-0, Library of Congress ref: "QA 279.5 Ber"

CLIFF1943: Clifford, Lt A. H: "Home Waters Enigma", in "Enigma Series Vol. 8, Reports from England Comint Tech Paper TS-10/E-7": National Archives, College Park, Maryland, RG 38, Radio Intelligence Publications (RIPs), Box 172, RIP 610

ERSK1992: Ralph Erskine, "The German Naval Grid in World War II", Cryptologia, 16 (1992) pp41-51.

HINS1993: Hinsley and Stripp (ed): "Codebreakers - the inside story of Bletchley Park", Oxford University Press 1993, ISBN 0-19-285304-X

KAHN1996: Kahn, D: "Seizing the Enigma", Arrow Press 1996, ISBN 0-09-978411-4

LEE1989: Lee, Peter. M: "Bayesian Statistics - An Introduction", Hodder Press 1989, ISBN 0-340-67785-6, Library of Congress ref: "QA 279.5 Lee2".

MAHON1947: Mahon, A. P: "The History of Hut 8, 1939-1945", available from the American National Archive and also from the Public Record Office, Kew, Surrey, HW 25/2.

SALE2000a: Sale, A: "Lecture Notes on Naval Enigma, Part 1", (see here) .

SALE2000b: Sale, A: "Lecture Notes on Naval Enigma, Part 2", (see here) .

USN1943: "U.S Navy Report on Banburismus", 1943 (proper citation needed).

WELCH1997: Welchman, G: "The Hut Six Story" (revised version), M & M Baldwin Press 1997, ISBN 0-947712-34-8

WHELA1944: Whelan, R. "The Use of Hollerith Equipment" NAUK, PRO HW 25/22, and in "Report on IBM Operations and Overseas Interception": NACP RG 457, HCC, Box CBTE 28, Nr. 3621).