# The Language's Impact on the Enigma Machine

Daniel Matyas Perendi and Prosanta Gope, *IEEE, Senior Member*

***Abstract*—The infamous Enigma machine was believed to be unbreakable before 1932 simply because of its variable settings and incredible complexity. However, people realised that there is a known pattern in the German messages, which then significantly reduced the number of possible settings and made the code breaker's job easier. Modern cryptanalysis techniques provide a lot more powerful way to break the Enigma cipher using letter frequencies and a concept called index of coincidence. In turn, this technique only works well for the English language(using the characters of the English alphabet), but what if we encountered an Enigma machine designed for the Hungarian language, where the alphabet consists of more than 26 characters? Experiments on the Enigma cipher with different languages have not been done to date, hence in this article we show the language's impact on both the machine and the cipher. Not only the Hungarian, but in fact, any language using more characters than the English language could have a significant effect on the Enigma machine and its complexity if there existed one. By a broad comparative analysis, it is proven that the size of the alphabet has a significant impact on the complexity and therefore the cryptanalysis.**

***Index Terms*—Hungarian; Enigma; cryptanalysis; cipher; complexity**

## I. Introduction

THIS investigation revolves around the famous Enigma machine, which was used by the German military to encrypt and protect commercial, diplomatic and military communication during World War II. Especially during the war, the need for secrecy was larger than ever, hence the Enigma was a lifeline for the German army as it provided highly complicated and secure encryption. This machine contained a series of interchangeable rotors, which rotated every time a key was pressed to keep the cipher changing continuously. This was combined with a plugboard on the front of the machine, where pairs of letters were transposed; these two systems combined offered approximately 107,458,687,327,250,619,360,000 (107 sextillion) possible settings to choose from, which back then seemed unbreakable. This "unbreakable" machine was broken by the gigantic effort of the British in Bletchley Park with Alan Turing's involvement in January 1940. The Enigma Machine relied on one default alphabet, which raises the question: What if this machine existed for different alphabets of different languages? The aim of this project is to experiment with different languages and alphabets to observe the effect they have on the machine and the cipher.

Contact: Daniel Matyas Perendi, Email: perendi.matyas98@gmail.com

### A. General history of the Enigma

Fig. 1: The Enigma machine (Rijmenants 2004)

As Figure 1 shows, the Enigma machine which is looked very much like an old typewriter with some extra elements. This amazing piece of engineering was mainly used in World War II by the German forces to securely transmit sensitive information across the battlefield. The first cipher machine, Enigma A appeared in 1923. Its successor, Enigma B was introduced soon after Enigma A, however these machines were extremely heavy (around 50kg) and quite big in size, so they weren't suitable for military usage where portability was key. A few years later Enigma C was equipped with the reflector and the lampboard replaced the "printer" part. This solution was a lot more compact and hence more suitable for military use. Enigma D was introduced in 1927 and appeared in several different versions with different rotor configurations across Europe. This machine had three rotors which could be set in one of the 26 positions(one for each character of the alphabet). In 1932, the Wehrmacht(aka. Enigma I) replaced the commercial Enigma D and extended that with the plugboard, which was attached to the front of the machine. The plugboard created a huge number of extra setting possibilities, hence this component became the target of cryptographic attacks. This version was introduced on a larger scale in the Heer(Army) and the Luftwaffe(Air Force). The German Navy also involved this machine in their communication protocols, however they extended the set of rotors to 8 and they named it "M3". Although they thought this machine was unbreakable, an

admiral called Karl Dönitz insisted on adding an extra rotor for greater security(Figure 2). This version was named M4. (Rijmenants 2004)

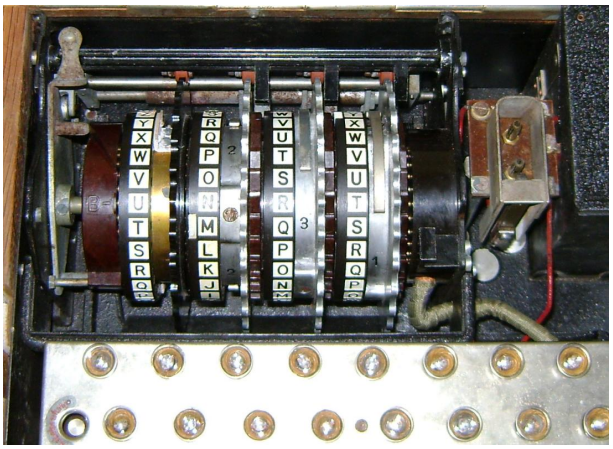

Fig. 2: The Enigma M4 with open cover (Rijmenants 2004)

## II. Related works

This section introduces some of the most important Enigma-breaking approaches that have significantly evolved over the years. We will go all the way back to the very first successful attempt, which will be followed by the famous Bletchley Park effort and finally wrapped up by the most recent and modern technique.

### A. Polish mathematicians

According to Gaj & Orłowski (2003) and Rijmenants (2004), the very first successful effort was made by a group of Polish mathematicians (Marian Rejewski, Jerzy Róycki, Henryk Zygalski), which focused on the beginnings of the German messages, because back in 1932 they all started with a 6-letter sequence, which essentially contained the key to the system. This 6-letter sequence was constructed from two successive encryption of three letters. These letters were unique for each message, however they were all encrypted with the actual daily settings of the Enigma machine. This led to significant information leakage, as the letters in position 1 and position 4, 2-5, 3-6 were the same before the encryption took place. Using this information as a basis, they could recover the missing permutations (Borowska & Rzeszutko (2014)). This group also invented an electro-mechanical machine (the Bomba), which sped up the breaking process. This method only worked until 1939, when the cipher design changed and the 6-letter sequence was eliminated.

### B. Bletchley Park

Codebreakers : the inside story of Bletchley Park presents the inside story of Bletchley Park and its importance in winning World War II. Numerous different signals(German, Japanese and Italian) were successfully intercepted and broken here, which provided an enormous amount of help to the Allied commanders on different fronts of the battlefield. Many of these messages were encrypted with different “versions” of the Enigma machine, which made the code-breakers’ job a lot harder. Bletchley Park was divided into smaller teams(Huts), where each team focused on different areas of the deciphering procedure. Hut 6 carried out the breaking of the three-wheel Enigma, Hut 8 dealt with the naval four-rotor Enigma, Hut 3 and Hut 4 respectively were in charge of producing and transmitting valuable intelligence -sourced from the deciphered messages produced by Hut 6 and 8- to the competent authorities. Various great minds contributed to Bletchley Park’s success: Alan Turing, Hugh Alexander, Gordon Welchman, Dilly Knox and many others. Different machines required different code-breaking techniques(Section II-C): Bombe machines, and the use of cribs and menus.

### C. The Bombe

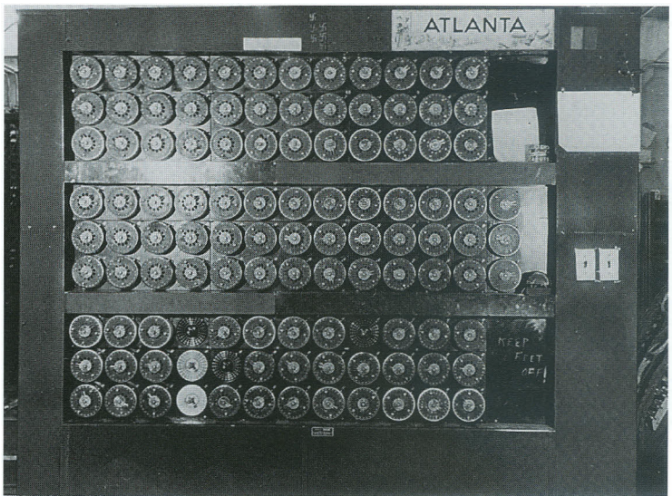


Fig. 3: The Bombe machine (Gladwin 1997, p. 211)

The previously mentioned Polish Bomba machine lost its usefulness due to German procedural changes (1939-1940), which prompted Alan Turing to design his version of this great code-breaking machine(Copeland 2020). Alan Turing started working on his machine in 1939 with more or less success -due to the complexity of the Enigma machine- and finished it in 1940 with an important change -the diagonal board- proposed by Gordon Welchman. The ground principle of this electro-mechanical instrument is to discover the daily Enigma key by testing all the possible settings, however the time required to exhaust all possibilities at first exceeded the 24 hours at disposal. In order to break the code within the 24-hour window, some changes had to be done. Cribs and menus combined with the Bombe resulted in a short enough running time of the searching method. Cribs are known pieces of a message, which were repeatedly used by the Germans during the war such as ”Wettervorhersage” meaning weather forecast and ”Keine besonderen Ereignisse” meaning nothing special to report. We know that no letter can be encrypted to themselves, so one could align the crib with the ciphertext the way it is shown on Figure 4.

After a valid alignment, a menu could be created from the letter connections (Figure 5), which then was plugged into the back of the Bombe machine as an electric circuit.

| ... | J | X | A | T | G | B | G | G | Y | W | C | R | Y | B | G | D | T | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | E | T | T | E | R | B | E | R | I | C | H | T | | | | | |

| ... | J | X | A | T | G | B | G | G | Y | W | C | R | Y | B | G | D | T | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | W | E | T | T | E | R | B | E | R | I | C | H | T | | | | |

| ... | J | X | A | T | G | B | G | G | Y | W | C | R | Y | B | G | D | T | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | W | E | T | T | E | R | B | E | R | I | C | H | T | | | |

Fig. 4: A crib

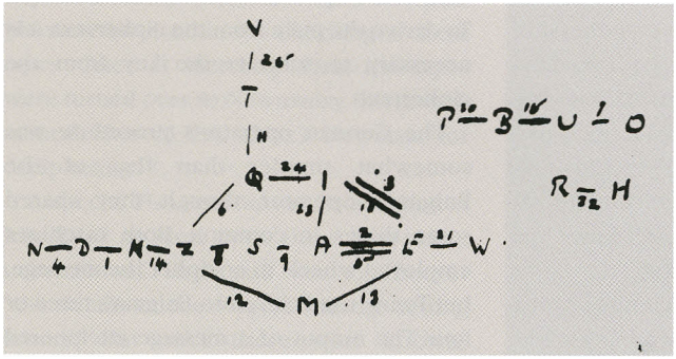

Fig. 5: A menu (Gladwin 1997, p. 211)

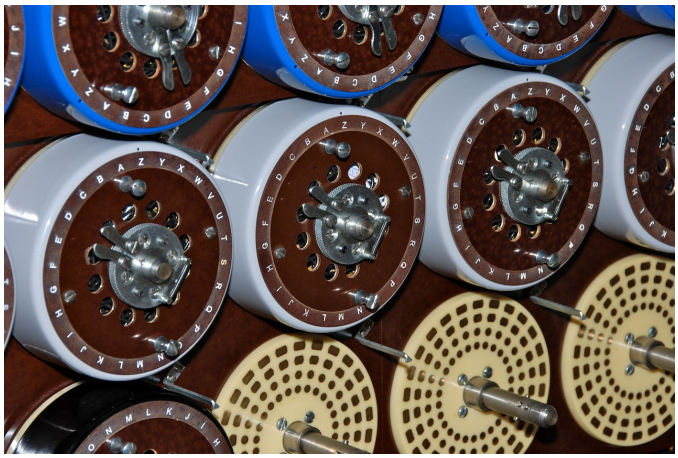

Fig. 6: The drums (CryptoMuseum 2009)

The Bombe can be thought of as 36 interconnected Enigma machines, where one drum represented one rotor and every drum rotated synchronously through all the $26^3$ possibilities (Figure 6). The front of the machine was responsible for working through all the 17,576 different rotor positions and stopping upon finding the correct settings (rotor order, rotor positions and plugboard connections) (Carter 2010). This "Stop" was the moment of relief, because the code-breakers knew that at that moment they found the daily key for the Enigma machine, so they could decipher all of that day's intercepted messages.

### *D. Modern approaches*

The most recent effort(Ostwald & Weierud (2017)) makes use of Friedman's idea of index of coincidence, which in simple terms is a measure of letter distributions in the candidate text Friedman (1922). Index of coincidence (also referred to as IoC or IC) calculates the probability of selecting two matching letters at random from a given text. This is useful, because letter distributions in natural languages are not even, hence the basic idea is to match the decryption attempt's IC to the language's IC. The index of coincidence can be calculated by using the following equation(*Index of Coincidence* (n.d.)):

$$IC = \frac{1}{N(N-1)} \sum_{i=1}^{n} F_i(F_i - 1) \qquad (1)$$

where

| | |
|---|---|
| N | is the length of the text |
| n | is the length of the alphabet |
| $F_i$ | is the occurrences of the *i*th letter of the alphabet |

If *N* is large enough i.e. *N* approaches infinity, we can calculate the expected IC for the language itself:

$$IC_{language} \approx \sum_{i=1}^{n} p_i^2 \tag{2}$$

where $p_i = \frac{F_i}{N}$. Using IC calculations as "validation" methods with hill climbing(Hillclimbing the Enigma Machine from Sullivan & Weierud (2006)) simplifies the brute-force technique significantly. Note that for a new wire, the IC is calculated for every possible rotor setting! This method works fine for the first few correct wires, but unfortunately, it fails to find the rest as it is described in "Modern breaking of Enigma ciphertexts"(Ostwald & Weierud 2017, p. 403-409). Bigram and trigram scores have been proven to be useful in finding the remaining wires. As amazing as it sounds, this method is not entirely robust either as its efficiency depends on the length of the text(we use this number to calculate the IC score), the shorter the text is, the more difficult/incorrect this approach is.

## III. Motivation

As discussed in the Section II-D, for breaking messages of the 26-letter general English alphabet there exist some very efficient modern code-breaking techniques, but how these techniques are affected if an Enigma machine is suited for another language -using a larger letter set- existed? Such experiments haven't been done to date, therefore it is the perfect opportunity to investigate the impact of a different alphabet on the complexity and the structure of the Enigma. The authors have a strong Hungarian background, hence it is worth starting the analysis with the Hungarian language, which could give a strong starting point for further -more general- analysis. The main objective is to observe the alphabet's influence on the Enigma cipher. All the techniques that have been used previously in the breaking process heavily rely on the complexity of this brilliant machine, hence observing the patterns in the increasing complexity would prove the cipher's greater security. In order to start this process, we have to go back to the fundamental complexity calculations to convert them into a parametric format, which will apply to any language and its alphabet.

## IV. Evaluation process

The above-mentioned variants of the Enigma machine have been broken by either mathematicians or cryptanalysts thanks to their great efforts. However, all of their solutions are built upon the ground complexity of the original machine. Nowadays people use numerous languages all over the world, which fuelled the idea of tailoring the Enigma machine to different languages. The base case of this experiment is the Wehrmacht machine, which will be used as the basis of comparisons. It is suspected that the size of the letter set(alphabet) is directly proportional to the complexity, therefore this is the main hypothesis that requires further evidence. The Hungarian language will be tested first, which will be followed by a general solution that can be applied to any languages that use a different alphabet.

## V. Analysis and result

### A. The Hungarian language

| A | Á | B | C | CS | DZ | DZS | E | É | F | G | GY | H | I | Í | J | K | L | M | N | NY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| O | Ó | Ö | Ő | P | Q | R | S | SZ | T | TY | U | Ú | Ü | Ű | V | W | X | Y | Z | ZS |

Fig. 7: The Hungarian 44-letter alphabet

The Hungarian language uses a very unique alphabet consisting of 44 letters (Figure 7), furthermore -as any other language- it also has its unique characteristics(special letter connections, words). Using this information it is possible to construct an Enigma machine using 44-letter rotors, an extended plugboard and a fixed 22-pair reflector. The complexity of this machine(assuming it uses 3 out of 5 rotors) can now be calculated the following way:

- Rotors: $(5 \times 4 \times 3) \times 44^3 \times 44^2 = 9,894,973,440$
- Plugboard using 18 wires: 39,282,388,067,747,317,859,706,965,625 (Figure 8)
- Total: 388,698,186,590,132,630,853,038,071,042,368,000,000

| | |
|---|---|
| 1 | 946 |
| 2 | 407,253 |
| 3 | 105,885,780 |
| 4 | 18,609,425,835 |
| 5 | 2,344,787,655,210 |
| 6 | 219,237,645,762,135 |
| 7 | 15,534,553,185,431,280 |
| 8 | 844,691,329,457,825,850 |
| 9 | 35,477,035,837,228,685,700 |
| 10 | 1,153,003,664,709,932,285,250 |
| 11 | 28,929,910,132,721,937,339,000 |
| 12 | 556,900,770,054,897,293,775,750 |
| 13 | 8,139,318,946,956,191,216,722,500 |
| 14 | 88,951,128,491,735,518,297,038,750 |
| 15 | 711,609,027,933,884,146,376,310,000 |
| 16 | 4,047,276,346,373,966,082,515,263,125 |
| 17 | 15,712,955,227,098,927,143,882,786,250 |
| 18 | 39,282,388,067,747,317,859,706,965,625 |
| 19 | 57,889,835,047,206,573,687,989,212,500 |
| 20 | 43,417,376,285,404,930,265,991,909,375 |
| 21 | 12,404,964,652,972,837,218,854,831,250 |
| 22 | 563,862,029,680,583,509,947,946,875 |

Fig. 8: Possible plugboard setting combinations for the 44-letter alphabet

The calculations suggest that this machine is approximately **3,617,187,183,818,893(3.6 quadrillion)**-times more complex than the Wehrmacht. The difference is colossal! There is however a major issue with using this alphabet the same way we would use the 26-letter English alphabet: This letter set contains double as well as triple-character letters (CS, DZ, DZS, GY, LY, NY, SZ, TY, ZS). This is a huge problem when it comes to the decryption process.

### B. The problem

During the encryption/decryption process the input text is read character by character, so it is a possibility that a single letter encrypts to a triple letter ( e.g. A → DZS), which not only causes a difference in the output length, but also affects at the decryption process (e.g. DZS → ?): There is no way we can tell whether these letters follow each other by coincidence, or they are meant to form this triple-character letter in this specific order. In the conventional process "D" will be pressed first on the machine, followed by "Z" and finally "S", which in the simplest case will produce an output of length 3 instead of the expected "A". This issue is demonstrated on Figure 9.

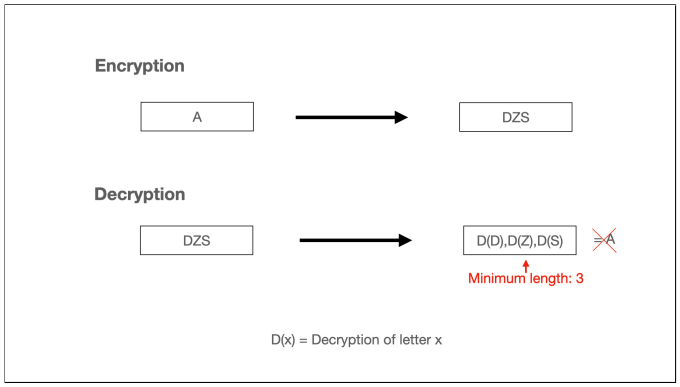


Fig. 9: The double and triple-character problem

### C. The solution

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Á | B | C | D | E | É | F | G | H | I | Í | J | K | L | M | N | O |
| Ó | Ö | Ő | P | Q | R | S | T | U | Ú | Ü | Ű | V | W | X | Y | Z | |

Fig. 10: The Hungarian 35-letter alphabet

The Hungarian alphabet contains 9 letters built up from either two or three characters. All of these characters can be constructed from the single characters of the alphabet, hence we are allowed use the alphabet without the non-singular letters, which will result in an alphabet of size 35(Figure 10). This change also has a great effect on the complexity of the machine and consequently on the cryptanalysis as well. Redoing the calculations with the adoption of the new table of possible plugboard combinations and the change of the rotors will result in a machine that is **42,094,345.5(42 million)**-times more complex than the original version. This number is still insanely huge, however it is nowhere near 3.6 quadrillion. Judging from these calculations a pattern seems to emerge: The larger the alphabet is, the more complex the Enigma machine gets, thus the longer it takes to break the code with either old or modern techniques. This matter is the subject of further investigation, which is conducted in the following section.

| | |
|---|---|
| 1 | 595 |
| 2 | 157,080 |
| 3 | 24,347,400 |
| 4 | 2,471,261,100 |
| 5 | 173,482,529,220 |
| 6 | 8,674,126,461,000 |
| 7 | 313,507,713,519,000 |
| 8 | 8,229,577,479,873,750 |
| 9 | 156,361,972,117,601,250 |
| 10 | 2,126,522,820,799,377,000 |
| 11 | 20,298,626,925,812,235,000 |
| 12 | 131,941,075,017,779,527,500 |
| 13 | 558,212,240,459,836,462,500 |
| 14 | 1,435,402,904,039,579,475,000 |
| 15 | 2,009,564,065,655,411,265,000 |
| 16 | 1,255,977,541,034,632,040,625 |
| 17 | 221,643,095,476,699,771,875 |

Fig. 11: Possible plugboard setting combinations for the 35-letter alphabet

The following table summarises the key differences between the English and Hungarian languages:

### D. Deriving a general formula

There are two main factors in the matter of the Enigma machine's complexity calculation: The number of possible rotor settings and the full range of combinations that the plugboard yields. The aim is to derive a parametric formula, which can show the effect of the parameter(number of extra letters in the alphabet) on the newly "constructed" Enigma machine's complexity. Before we dig into the calculations, it is important to lay down two ground rules to ensure the cipher's and the machine's correct mechanisms:

- The alphabet can only contain single characters
- The number of wires used for the plugboard is at most half of the alphabet's size

*1) The Rotors:* For the purpose of this experiment a 3-rotor Enigma is considered, where 3 rotors are chosen from a total set of 5 rotors. The selection process yields $5 \times 4 \times 3 = \mathbf{60}$ combinations, which will remain constant in our formula. The variable part includes the rotor positions and the notch(ring) settings; Both of these depend on the size of the alphabet. In the original machine's case these equal to $26^3$ and $26^2$ respectively, so merging these two terms will result in $\mathbf{26^5}$. As these depend on the alphabet, we will add the parameter into the equation: $(26+x)^3 \times (26+x)^2$, where "x" is the number of extra letters compared to the English alphabet. Following mathematical transformations the final result is:

$$26^5 + [x^5 + 130x^4 + 6760x^3 + (10 \times 26^3)x^2 + (5 \times 26^4)x]$$

The expression in square brackets calculates the number of added possibilities. For example, 1 extra letter will add $1 + 130 + 6760 + (10 \times 26^3) + (5 \times 26^4) =$ **2,467,531** rotor settings.

| | English | Hungarian |
|---|---|---|
| Total number of letters | 26 | 35 |
| Rotor settings | 11,881,376 | 52,521,875 |
| Plugboard settings (second largest) | 150,738,274,900,000 | 1,435,402,904,039,579,475,000 |
| Expected IC | 0.0686 | 0.0549 |
| Total number of settings | 107,458,687,300,695,744,000,000 | 4,523,403,114,036,227,294,310,937,500,000 |

Fig. 12: Summative comparison of the two languages

*2) The Plugboard:* The plugboard provides an incredible amount of setting possibilities, therefore it is important to examine the equation and derive a formula where the number of wires and the number of the extra letters in the alphabet are the parameters. As previously mentioned, it is essential to keep the number of wires either at or below half of the alphabet's size, because the wires connect two letter-slots on the plugboard, hence the maximum number of wires the Enigma machine can handle is $\lfloor \text{alphabet size}/2 \rfloor$. The Wehrmacht's plugboard settings are calculated the following way in relation to the number of wires("n"):

$$\frac{26!}{n! \times (26-2n)! \times 2^n}$$

The next step is to introduce the second parameter: The number of extra letters in the alphabet("x"):

$$\frac{(26+x)!}{n! \times (26+x-2n)! \times 2^n}$$

This formula is lot more complicated than the one for calculating the added rotor complexity, therefore we will consider the numerator and the denominator separately. The numerator can be broken down into a multiplication:

$$26! \times \frac{(26+x)!}{26!}$$

and similarly the denominator:

$$n! \times (26-2n)! \times \frac{(26+x-2n)!}{(26-2n)!} \times 2^n$$

These two expressions are very similar to the original formula, thereby with some rearrangements we get the following result:

$$\frac{\frac{(26+x)!}{26!}}{\frac{(26+x-2n)!}{(26-2n)!}} \times \frac{26!}{n! \times (26-2n)! \times 2^n}$$

$$= [\frac{(26+x)! \times (26-2n)!}{26! \times (26+x-2n)!}] \times \frac{26!}{n! \times (26-2n)! \times 2^n}$$

where the expression within square brackets is the plugboard's complexity multiplier for a given number of wires("n") and a given number of extra letters("x").

### *E. The Final Formula*

In order to calculate the total difference in terms of complexity, the added rotor complexity and the plugboard multiplier has to be plugged into the general formula (**Rotor combinations(60) x Rotor settings x Plugboard settings**):

$$60 \times (26^5 + [x^5 + 130x^4 + 6760x^3 + (10 \times 26^3)x^2 + (5 \times 26^4)x]) \times [\frac{(26+x)! \times (26-2n)!}{26! \times (26+x-2n)!}] \times \frac{26!}{n! \times (26-2n)! \times 2^n}$$

Although this formula looks hectic and confusing, it is possible to make it look more pleasant. The previously derived formulas can be represented as symbols: the additional rotor settings are denoted as "***RA***" and the plugboard multiplier is indicated by "***PM***":

$$60 \times (26^5 + RA) \times (\frac{26!}{n! \times (26-2n)! \times 2^n} \times PM)$$

It can be rearranged further in such a way that the original Enigma machine's formula is multiplied by a number:

$$60 \times 26^5 \times \frac{26!}{n! \times (26-2n)! \times 2^n} \times [PM + \frac{PM \times RA}{26^5}]$$

$$= \text{Original formula} \times [PM + \frac{PM \times RA}{26^5}]$$

Taking everything into account, the expression in square brackets is what determines "how many times more settings the new machine has", or in other words, how many times more complicated the second machine is in relation to the number of wires used and the number of extra letters in the alphabet. As an example, an Enigma machine designed for one extra letter in the alphabet while using the default 10 wires has ∼4.66 times more setting possibilities than the Wehrmacht Enigma. However, this formula only calculates this multiplier accurately for "n" values between 0 and 13 as the original formula would produce a negative result for any larger "n" values. Although for larger "n", we can use the following formula to calculate the new number of possible settings, though can't compare it to the Wehrmacht Enigma for the previously mentioned reason:

$$60 \times (26+x)^5 \times \frac{(26+x)!}{n! \times (26+x-2n)! \times 2^n}$$

## VI. Conclusion

Based on the Wehrmacht Enigma machine's complexity calculations, a general formula had been derived for the purpose of proving the alphabet's influence on both the machine and the cipher. Since both known-plaintext attacks and ciphertext-only attacks depend on the entire key-space, the alphabet affects the machine's cryptanalysis likewise. It has also been proven that only a single extra letter increases the number of rotor combinations by 20 percent, the plugboard combinations by around 3.85 times and the total machine complexity by approximately 4.66 times. This difference expressed in terms of numbers yields a new total of **500,757,482,821,242,167,040,000** possible settings.

## References


Borowska, A. & Rzeszutko, E. (2014), 'The cryptanalysis of the enigma cipher. the plugboard and the cryptologic bomb.', *Computer Science* **15**.

Carter, F. (2010), 'The turing bombe', *The Rutherford Journal* **3**.

Copeland, B. J. (2020), 'Alan turing', *Encyclopedia Britannica* .

CryptoMuseum (2009), 'Crypto and cipher machines', https://www.cryptomuseum.com/crypto/index.htm (last accessed: 30.06.2021).

Friedman, W. (1922), 'The index of coincidence and its applications in cryptanalysis'.

Gaj, K. & Orłowski, A. (2003), Facts and myths of enigma: Breaking stereotypes, Vol. 2656, pp. 106–122.

Gladwin, L. A. (1997), 'Alan turing, enigma, and the breaking of german machine ciphers in world war ii', *Prologue* .

*Index of Coincidence* (n.d.), https://pages.mtu.edu/~shene/NSF-4/Tutorial/VIG/Vig-IOC.html (last accessed: 25.04.2021).

Ostwald, O. & Weierud, F. (2017), 'Modern breaking of enigma ciphertexts', *Cryptologia* **41**(5), 395–421.

Rijmenants, D. (2004), 'Cipher machines and cryptology', http://users.telenet.be/d.rijmenants/en/enigmamenu.htm (last accessed: 30.06.2021).

Sullivan, G. & Weierud, F. (2006), 'Hillclimbing the enigma machine'.