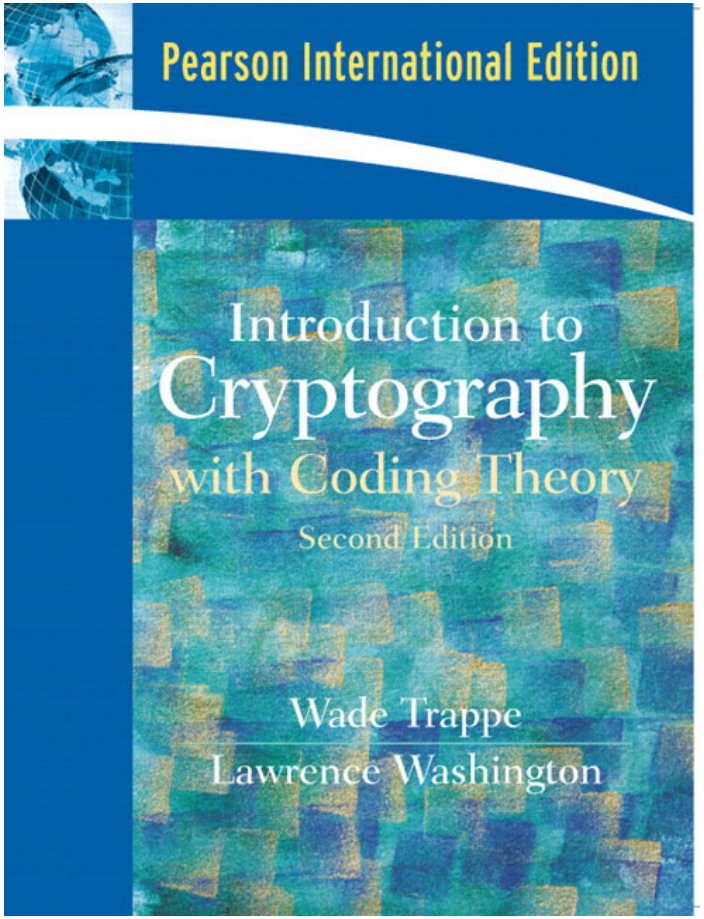
Cryptography Lecture 1
Principles and history

Course book, examination

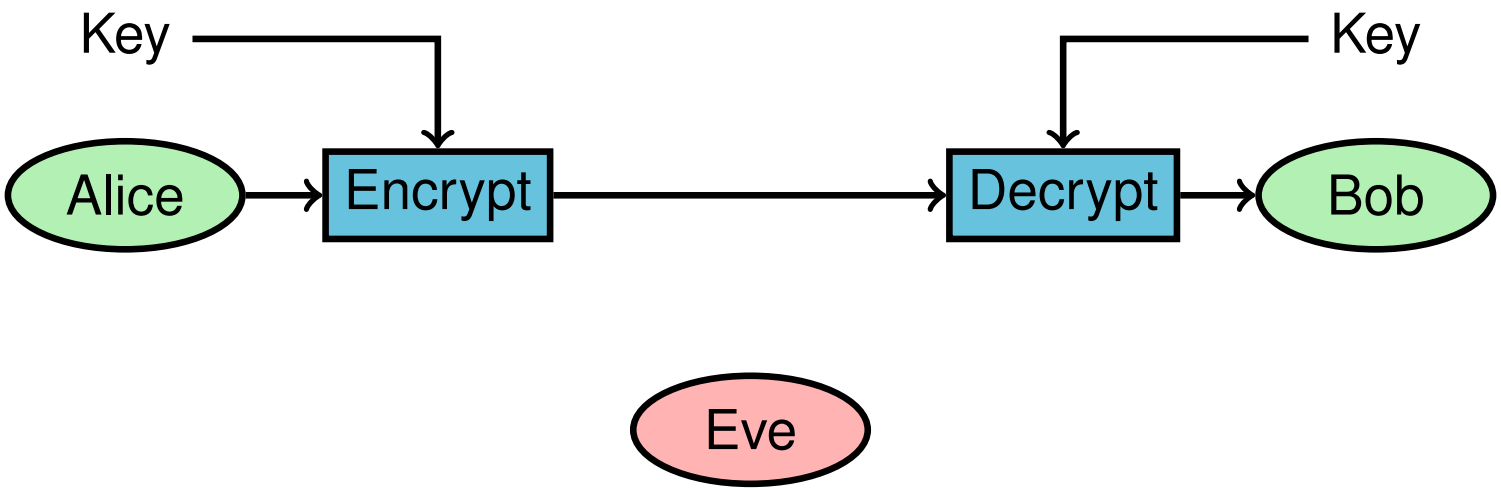
- 12 lectures
- 4 lab sessions
- Written exam

- You should register for lab 1 NOW
- Keep an eye out for instructions in lisam



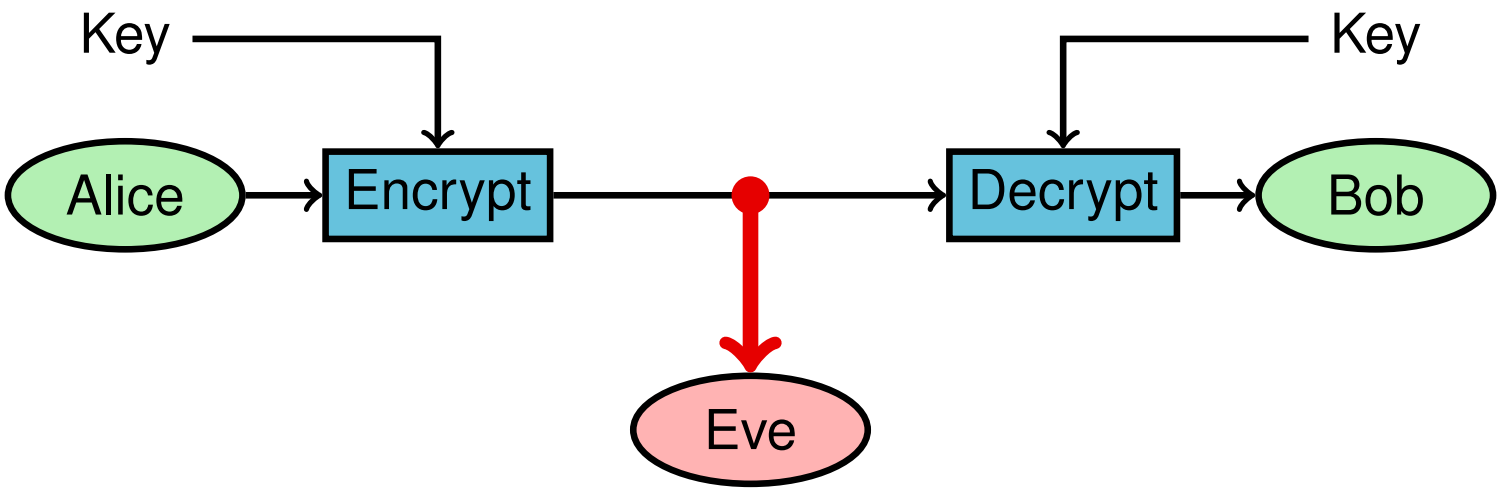
“Cryptography” is a Greek word that means “hidden writing”

Used to hide message from someone, and sometimes prevent them from creating a new message



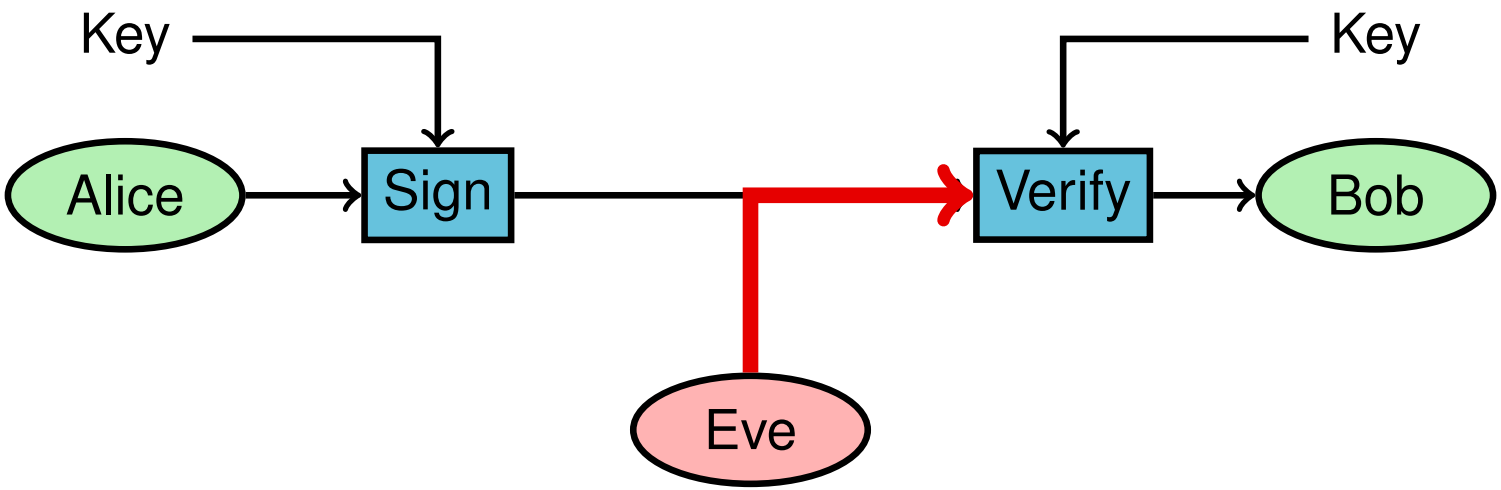
“Cryptography” is a Greek word that means “hidden writing”

Used to hide message from someone, and sometimes prevent them from creating a new message

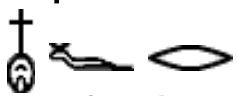


“Cryptography” is a Greek word that means “hidden writing”

Used to hide message from someone, and sometimes prevent them from creating a new message



The message is written using an alphabet in some language

- Egyptian hieroglyphs were unreadable until the Rosetta stone was found. This contained the same text in Ancient Egyptian hieroglyphs, in Demotic script, and in ancient Greek.
- For example, “Nefer” meaning “good”, “beautiful” could be written  or  or  or in a lot of other ways, like a picture of a horse
- Non-standard = Encrypted? Not really. . .

Terminology

- The *plaintext* is the information in its normal form
- The *ciphertext* or *cryptogram* is the transformed plaintext
- The secret parameter for the encryption (known only to the sender and intended recipients) is called the *key*
- The key decides how the transformation is done

Kerckhoff's principle

- A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

Terminology

- **Encryption** (Swedish **kryptering**) transforms a plaintext into a cryptogram
- **Decryption** (Swedish **dekryptering**) transforms a cryptogram back into plaintext using a known key
- **Cryptanalysis**, or breaking a cipher is transforming a cryptogram back to the original plaintext without previous knowledge of the key (Swedish **kryptoanalys**, **kryptoknäckning**, **forcering**)

The three basic types of cryptography

Steganography:	Disguise there is a message
Codes:	Look up in a secret table
Ciphers:	Use a general algorithm with a secret parameter known only to a select few

Steganography

- Not part of the course
- Oldest historic examples are writing on a slave's shaved head (no fast delivery needed, obviously) or on the wood beneath the wax of writing tablets.
- Other examples are writing with lemon juice, microdots, or using the least significant bits in digitally encoded pictures

- For example, used by Richelieu, who was rumored to like the "Cardan grille"

Codes

- Not the public codes treated in coding theory
- Tables list every possible plaintext for encryption and every possible ciphertext for decryption
- Listed items can be letters, sentences, names etc.
- Items not in the table are sent in clear

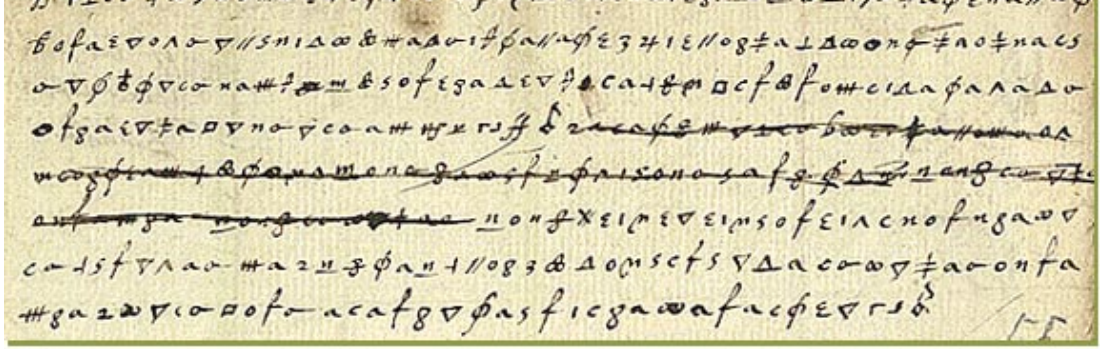
- For example, used by Mary Queen of Scots

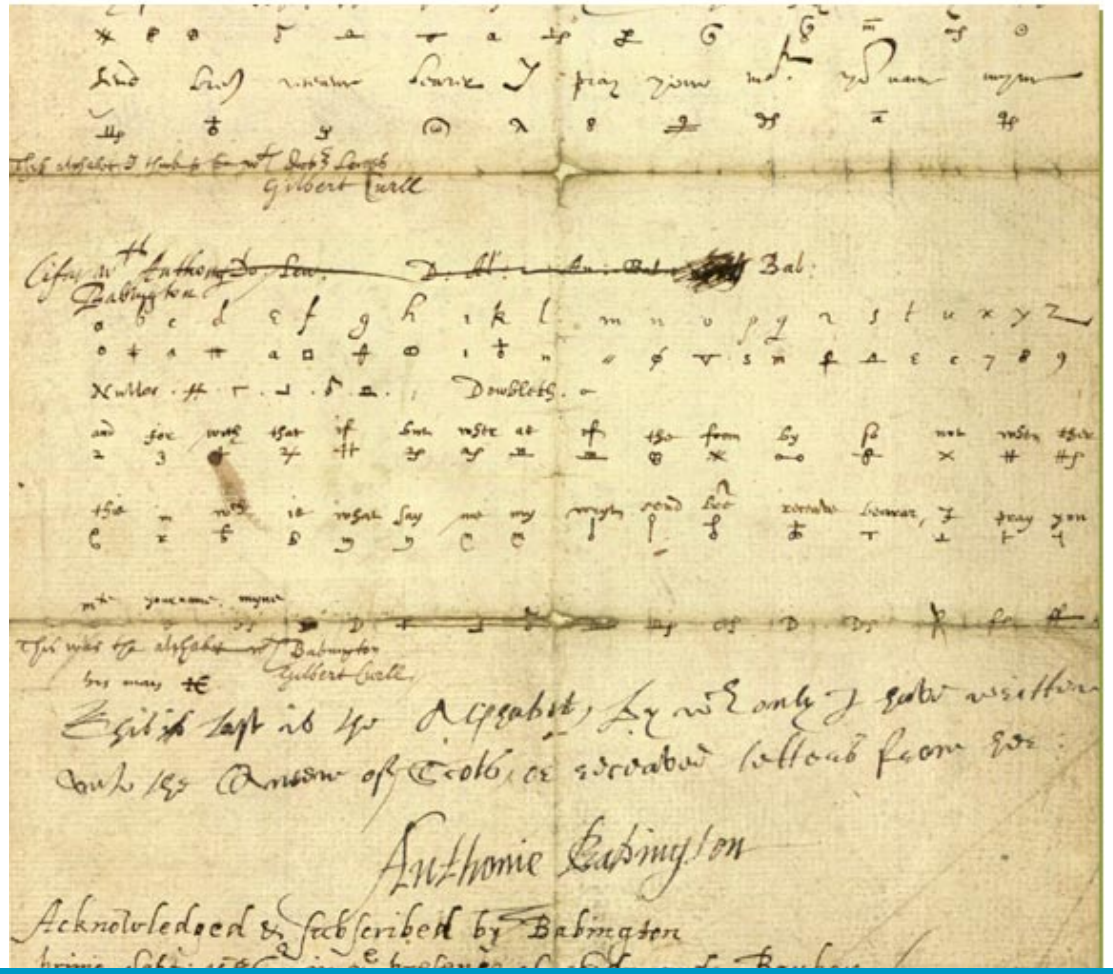
Trusting and breaking codes

- Code breakers used context, inference, pieces of corresponding plaintext and to some extent statistics to reconstruct codes.
- Mary, Queen of Scots, was prisoner at the mercy of her cousin Elisabeth of England.
- Mary's friends used codes in letters, because they (correctly) suspected that Elisabeth's agents might read them.
- Mary responded to a letter containing a plot to kill Queen Elisabeth, assuming that the received letter had not been read. Wrong assumption. . .

Code book

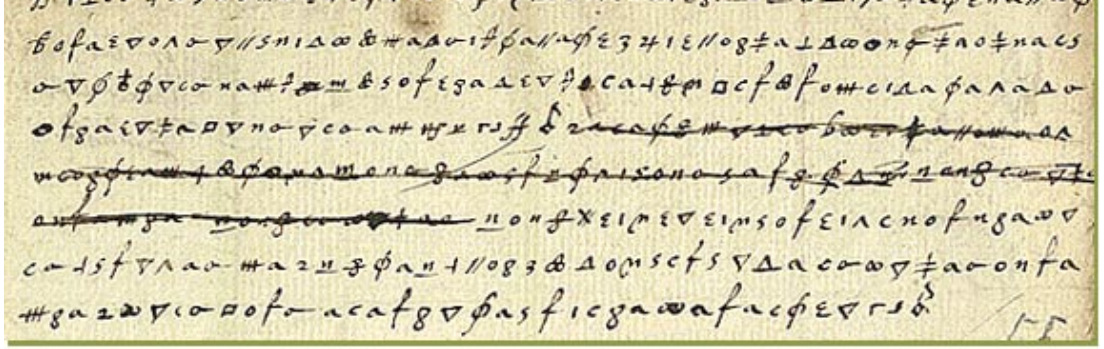
(Nomenclator)

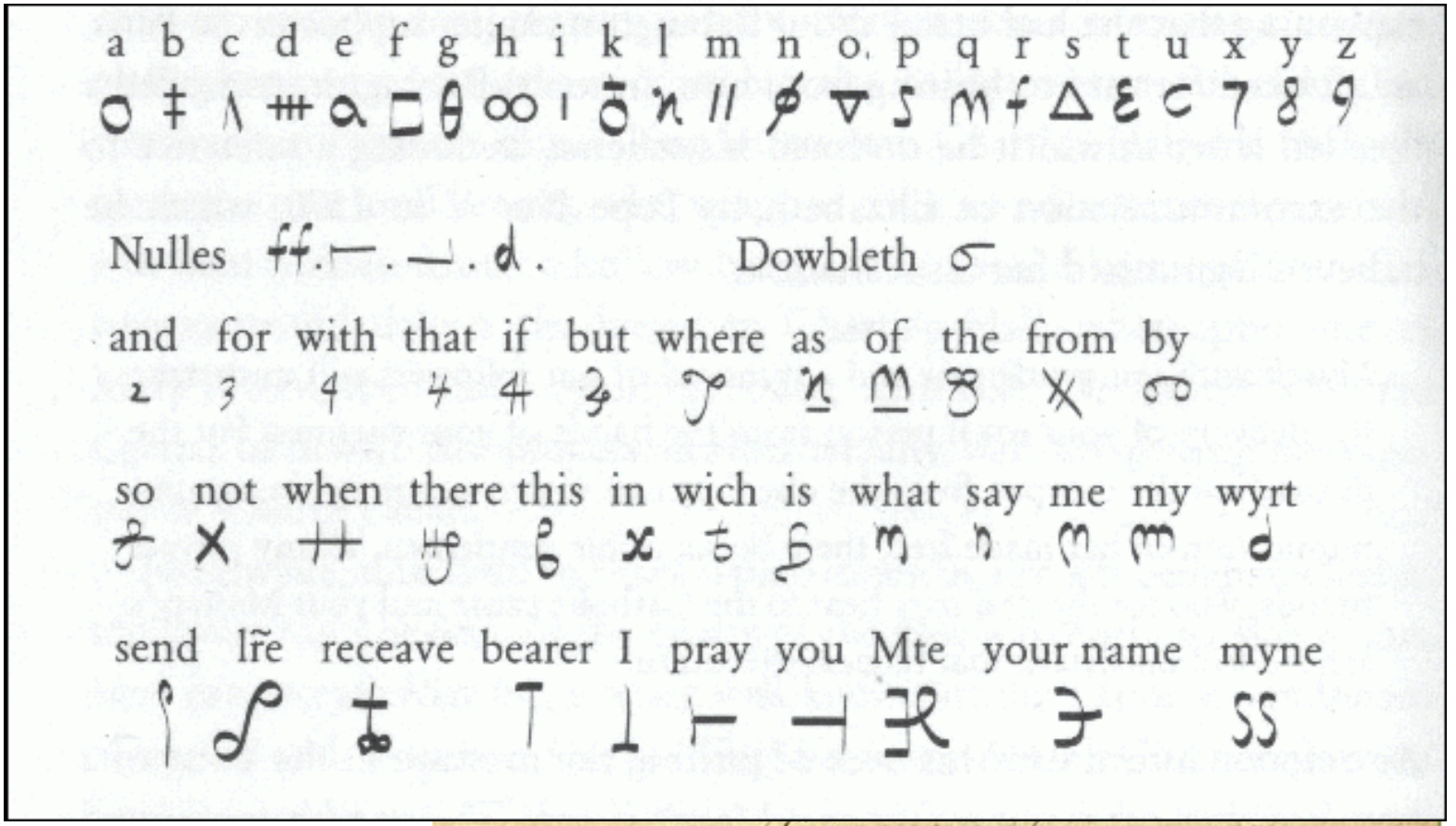


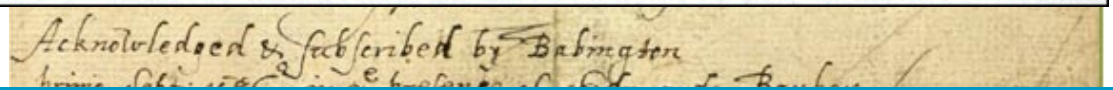


Code book

(Nomenclator)

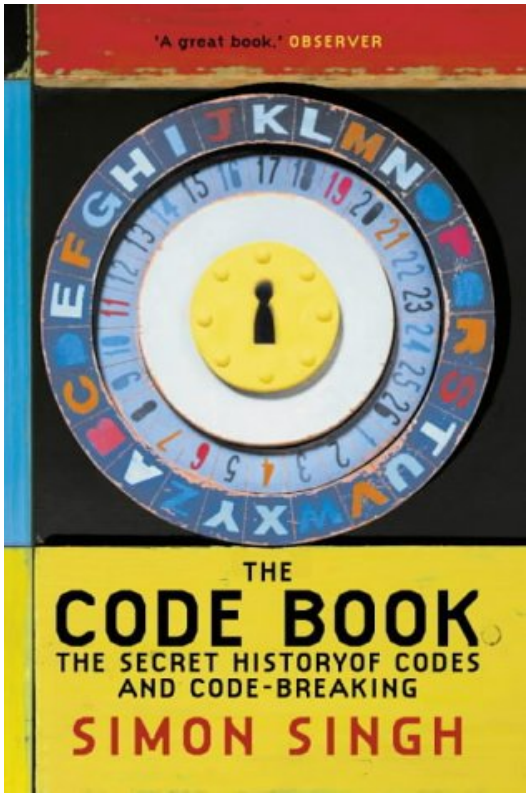






Another kind of code book

- Use an actual book
- Write your code as three numbers
- These might be (page, row, letter), or something else you have agreed on
- Decode by paging through the book



Ciphers (what we use in modern cryptography)

- Overlap as a method with code tables, when every possible plaintext is in the table.
- Use alphabets, which can consist of just printable characters, bit sequences of any fixed length or anything else defined as a finite set
- Plaintext and ciphertext do not necessarily use the same alphabet

Classical crypto 1: Skytale (Greek word)

- (NOT pronounced as English “sky tale”. Pronounce as in Swedish, German, Italian etc. with stress on the middle syllable)
- Wind a strip, one letter wide, as a tight spiral around a stick, write along the stick, unwind
- The key is the width of the stick.

How to break skytale

- Simply try different width of sticks, wind the strip around them so that letters are reasonably aligned with the central axis of the stick, and look out for the width that gives you a readable message.
- The slant of the letters gives a clear indication of the approximate width.

Skytale vs general transposition

- Transposition (permutation) ciphers use only the original plaintext letters, but write them in a different order. So skytale is one example.
- More common method: Write the plaintext, one letter per cell, in a table with fixed number of columns, rearrange the columns, and copy the letters row by row in the new order.
- Skytale can be done similarly, using a table with a fixed number of columns, reading out top to bottom instead of left to right.

w	r	i	t	e
t	h	e	p	l
a	i	n	t	e
x	t	o	n	e
l	e	t	t	e

e	r	w	t	i
l	h	t	p	e
e	i	a	t	n
e	t	x	n	o
e	e	l	t	t

How to break transposition

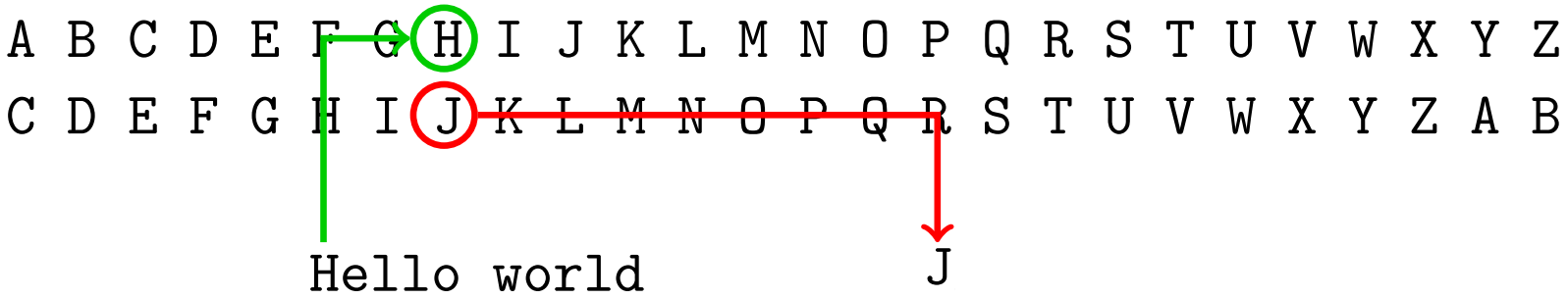
- Cryptogram is “toohwarkbeatnarspisonicto. . . ”
- Guess at the number of columns
- Search for the letters of common syllables or a known word on the same row, here: “tion”
- Rearrange columns so that the guessed syllable/word is formed and that the rest makes sense.

t	o	o	h	w
a	r	k	b	e
a	t	n	a	r
s	p	i	s	o
n	i	c	t	o

h	o	w	t	o
b	r	e	a	k
a	t	r	a	n
s	p	o	s	i
t	i	o	n	c

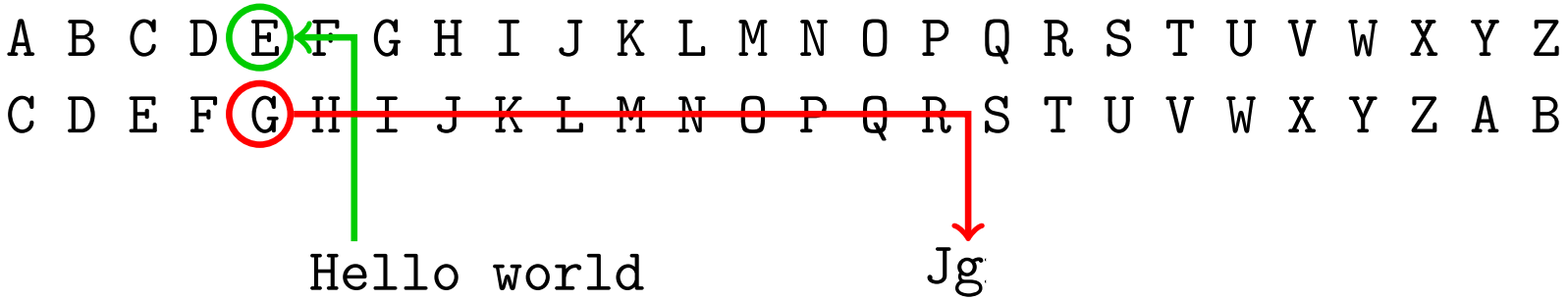
Classical crypto 2: Caesar cipher

- Exchange every plaintext letter into the letter *k* positions further on in the alphabet
- The key is the letter that A is transformed into



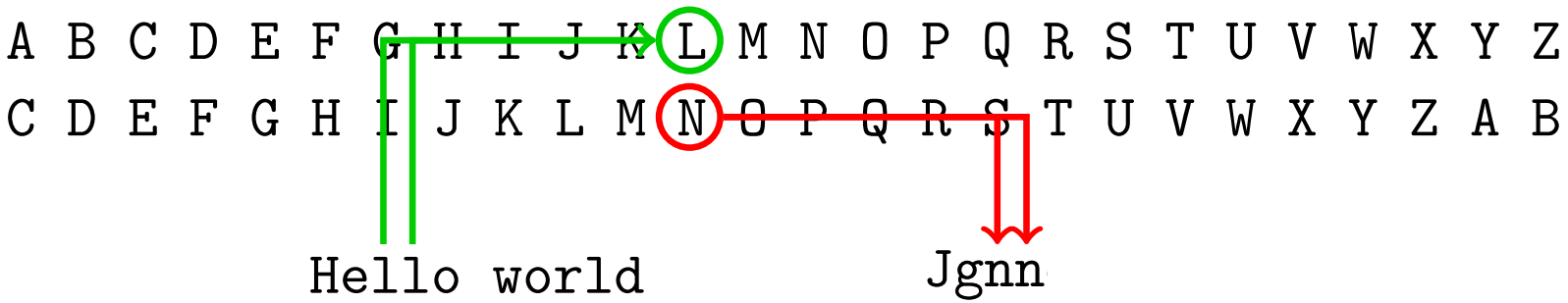
Classical crypto 2: Caesar cipher

- Exchange every plaintext letter into the letter *k* positions further on in the alphabet
- The key is the letter that A is transformed into



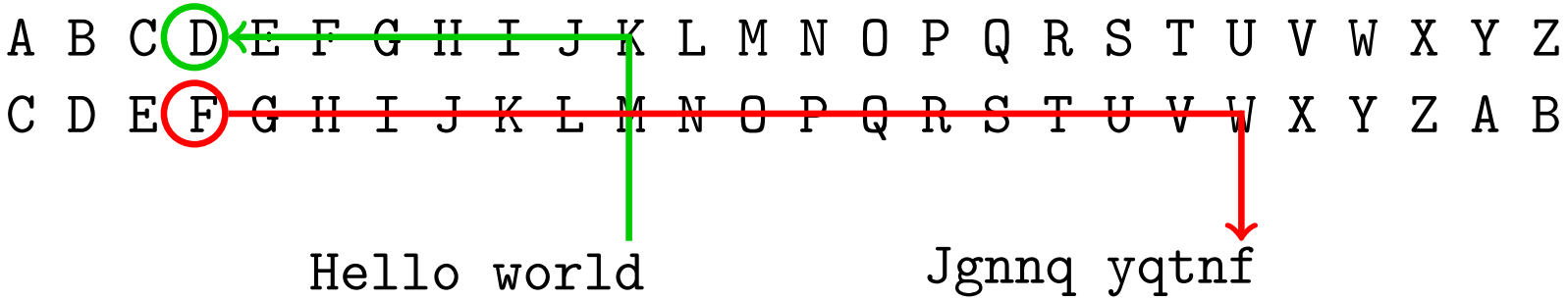
Classical crypto 2: Caesar cipher

- Exchange every plaintext letter into the letter *k* positions further on in the alphabet
- The key is the letter that A is transformed into



Classical crypto 2: Caesar cipher

- Exchange every plaintext letter into the letter *k* positions further on in the alphabet
- The key is the letter that A is transformed into



Alternative description of Caesar cipher

- Replace every plaintext letter with its (zero-offset) position in the alphabet (“A”=0, “B”=1, etc., up to the number of letters n)
- Express the key as an integer k using the same system
- If the plaintext as an integer is m, the cryptogram as an integer $c = m + k$ modulo n
- The cryptogram letter is then the letter corresponding to the number c

- The plaintext “H” gives $m = 7$, and $k = 2$ results in $c = 7 + 2 \pmod{26}$, so cryptogram is “J”

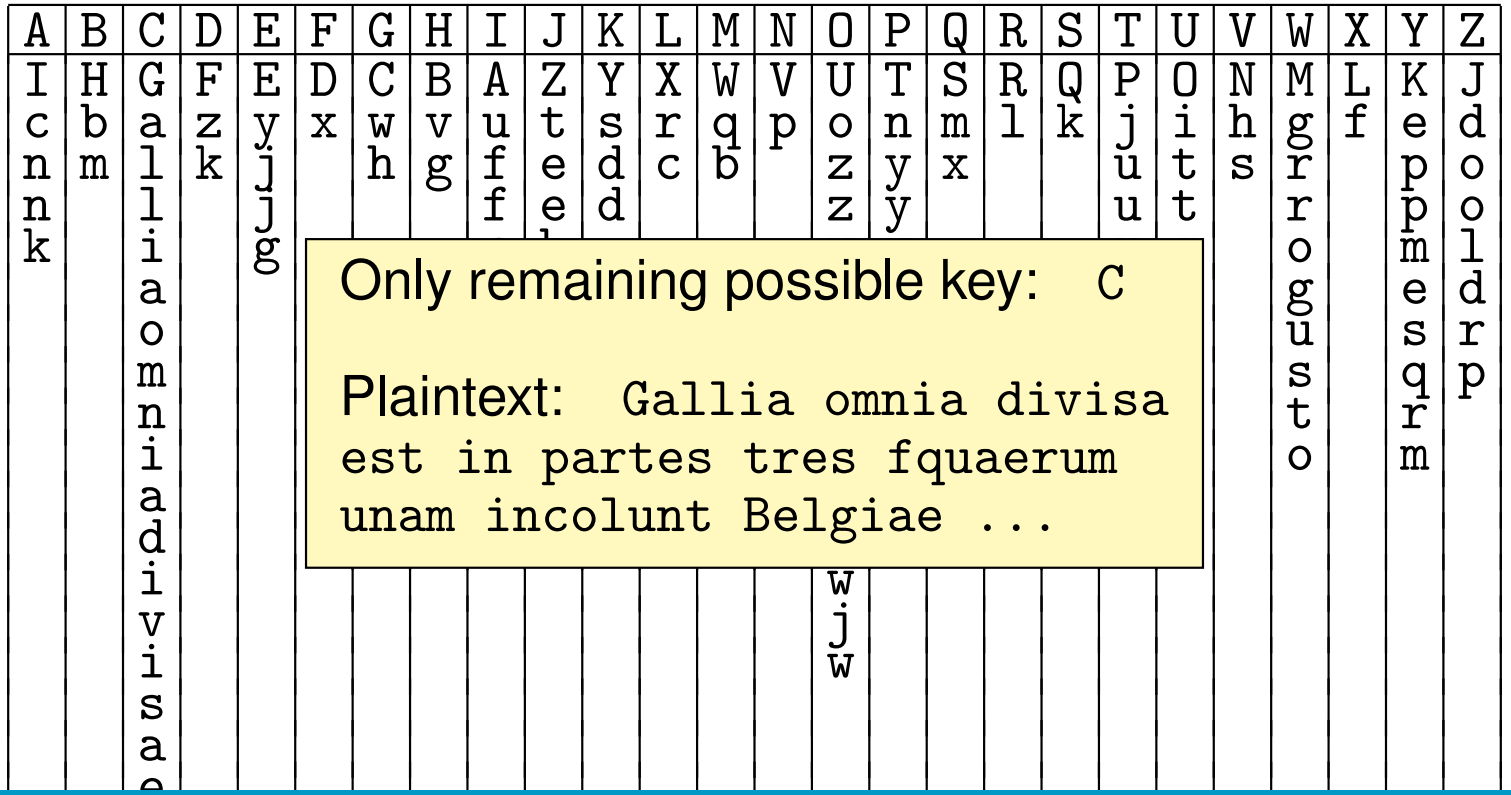
Breaking Caesar, example

- Cryptogram: Icnnkc qopkc fkxkuc guv kp rctvgu swcgtwo wpco kpeqnwpv Dgnikcg. . .
- Try each key, stop trying for each key when the plaintext becomes impossible

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Icnnk	Hbm	Gallia omniadivisae	Fzk	Eyjjg	Dx	Cwh	Bvg	Auffcuighc	Zteebthf	Ysdd	Xrc	Wqb	Vp	Uozzwocabworwjw	Tnyyvn	Smx	Rl	Qk	Pjuurjxv	Oittqiwuvq	Nhs	Mgrrogusto	Lf	Keppmesqrm	Jdoolderp

Breaking Caesar, example

- Cryptogram: Icnnkc qopkc fkxkuc guv kp rctvgu swcgtwo wpco kpeqnwpv Dgnikcg...
- Try each key, stop trying for each key when the plaintext becomes impossible



Runes and the Rök ciphers (800AD)

- Runes have a formal alphabet

ᚠᚢᚦᚨᚱᚲ ᚺᚾᛁᛃ ᛏᛒᛗᛚᛦ

f uthark hnias tbmlR

- The stone at Rök contains one line starting

ᛅᛁᚱᚠ ᛒᚠ ᚱᛒᚾᚺᚾ ᚠ ᛁᚾᛒᛅᚾᛏ ᚠ ᛅᚾᚺᚾᚢ

air f b f rbnhn f Inbant f anhnu

Substitute every rune with its neighbour to the right

ᛋᛅᚴᚢᛗ ᚢᚴᛗᛁᚾᛁ ᚢᛅᛁᛗ ᛋᛁ ᛒᚢᚱᛁ ᚾᛁᚦ

Sakum ukmini uaim si buri nith

The stone also contains numeric ciphers like “3,3; 3,2; 3,5” meaning “Thor”.

Strengthen the Caesar cipher

- The Caesar cipher is a simple shift $c = m + k$ modulo n
- The Affine cipher uses a second key integer j and encrypts using $c = jm + k$ modulo n
- Caesar has 26 possible keys, the affine cipher has $312 = 26 \cdot 12$ key values

Strengthen the Caesar cipher

- The Caesar cipher is a simple shift $c = m + k$ modulo n
- The Affine cipher uses a second key integer j and encrypts using $c = jm + k$ modulo n
- Caesar has 26 possible keys, the affine cipher has $312 = 26 \cdot 12$ key values
- The reason is that $j = 13$ cannot be used

Strengthen the Caesar cipher

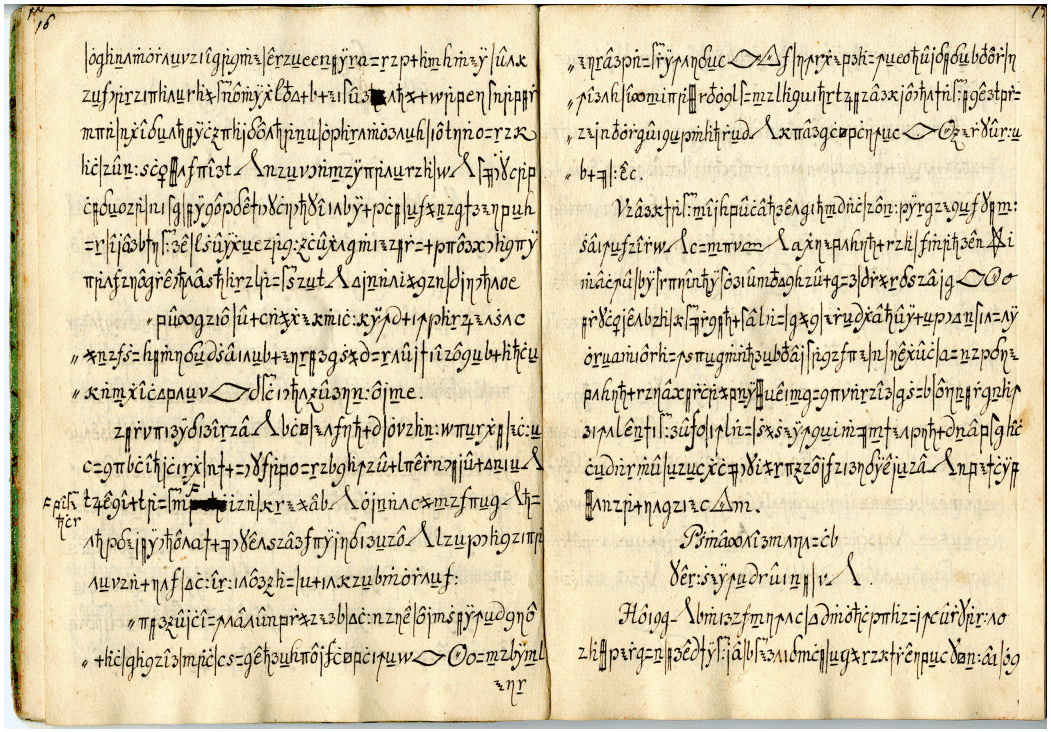
- The Caesar cipher is a simple shift $c = m + k$ modulo n
- The Affine cipher uses a second key integer j and encrypts using $c = jm + k$ modulo n
- Caesar has 26 possible keys, the affine cipher has $312 = 26 \cdot 12$ key values
- The reason is that $j = 13$ cannot be used
- With $j = 13$, the encryption will not be unique: even m will give $c = k \pmod{26}$, while odd m will give $c = 13 + k \pmod{26}$

Simple substitution (= code!)

- Create a table of plaintext characters and their corresponding crypto characters
- Crypto characters can be just ordinary letters, but also anything else.
- Each crypto character must occur only once in the table to enable unique decryption

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

Copiale

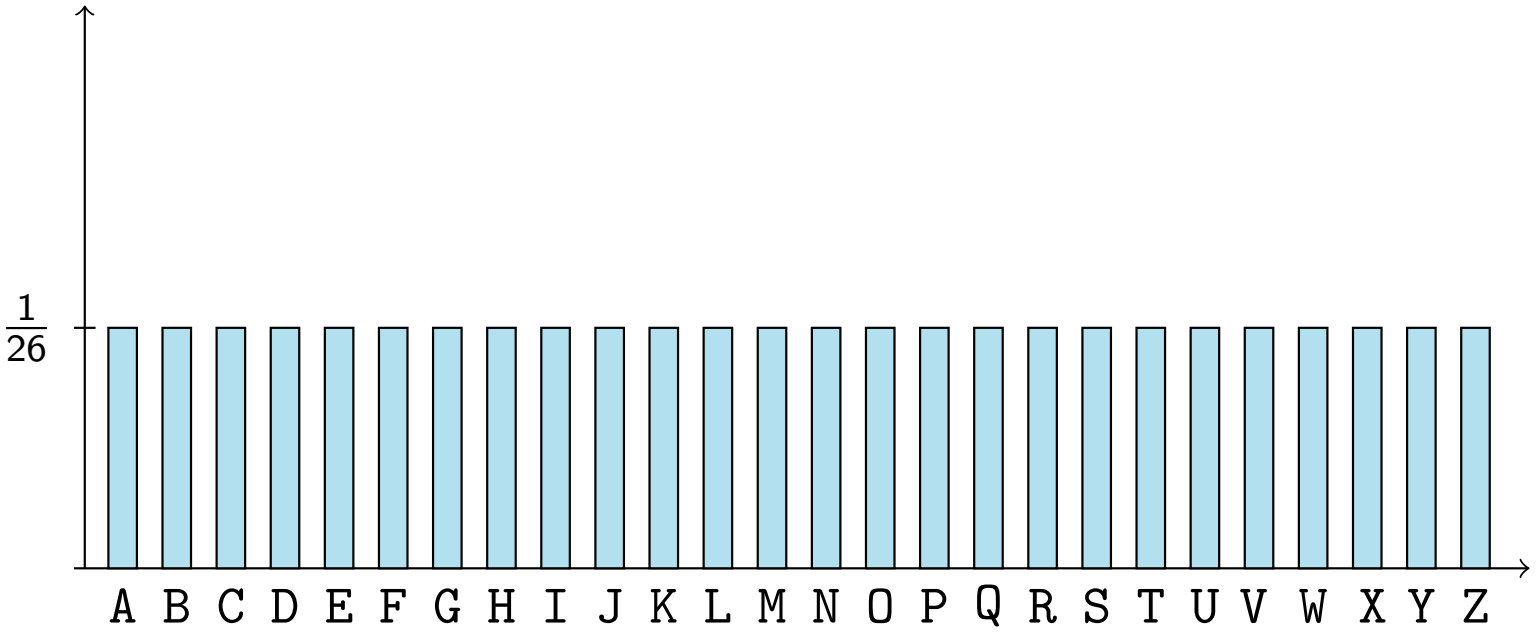


- Copiale is an encrypted manuscript from the late 1700's
- 75 000 handwritten characters
- Unbroken until 2011

Breaking simple substitution

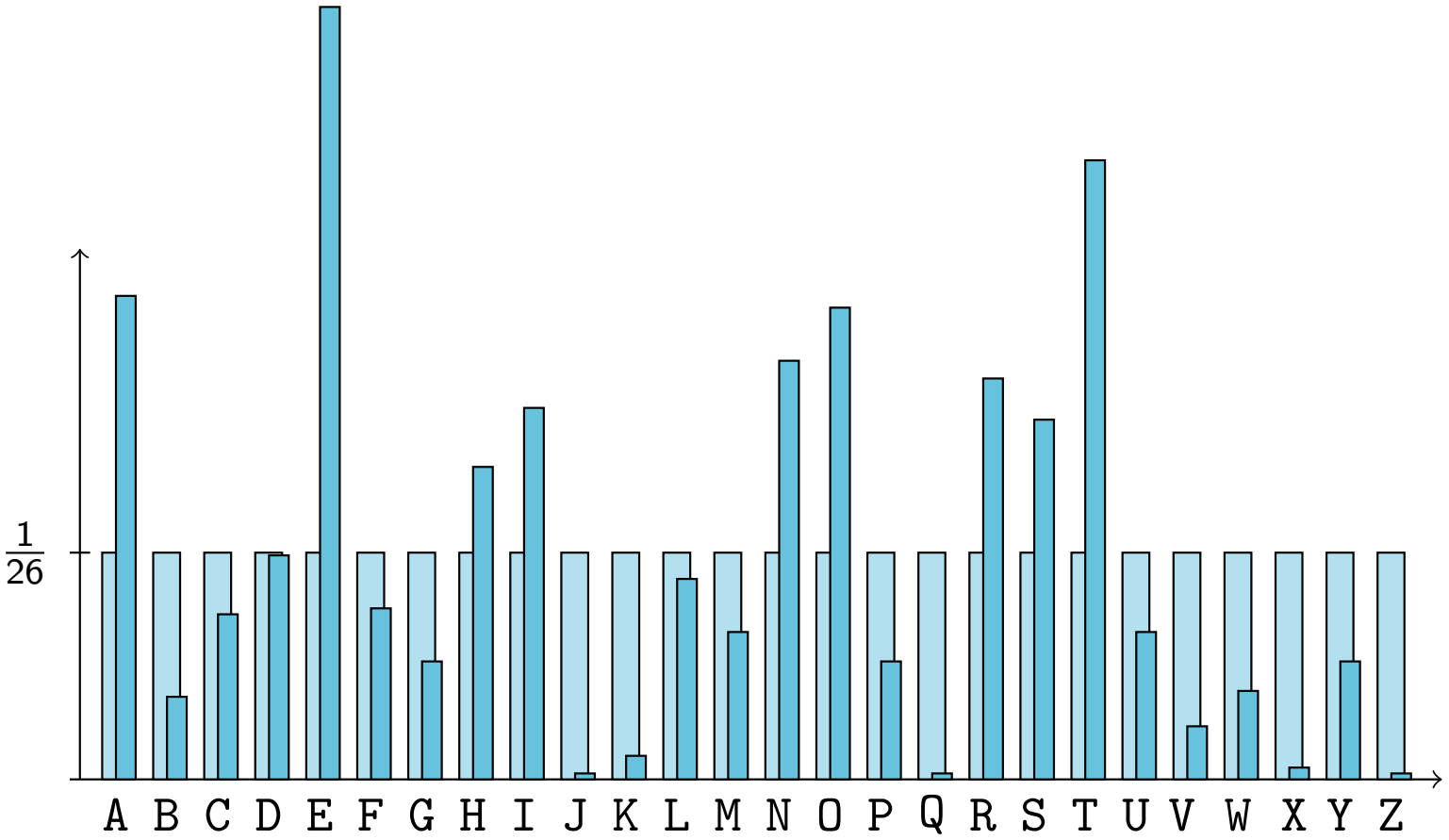
- Every crypto letter will occur exactly as often as its plaintext counterpart occurs in plaintext.
- Every combination of crypto letters (digrams, trigrams etc.) will occur as often as the corresponding plaintext combinations.
- Count how often letters, bigrams and trigrams occur in the cryptogram, and try to identify the ones corresponding to common plaintext letters and common letter combinations.
- Fill in so that remaining gaps form words.

The letter distribution of English is uneven



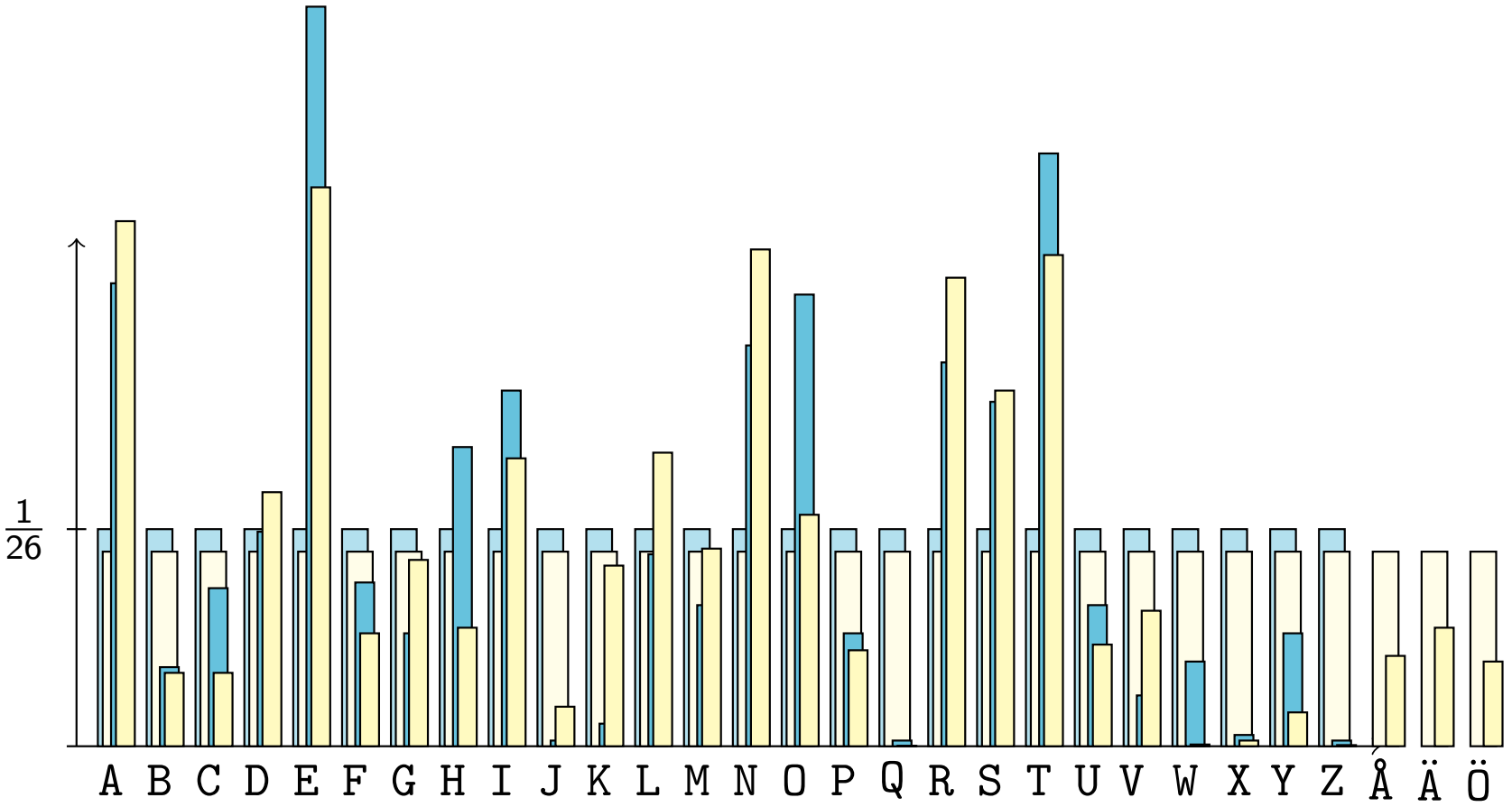
- An even distribution would look like the above

The letter distribution of English is uneven



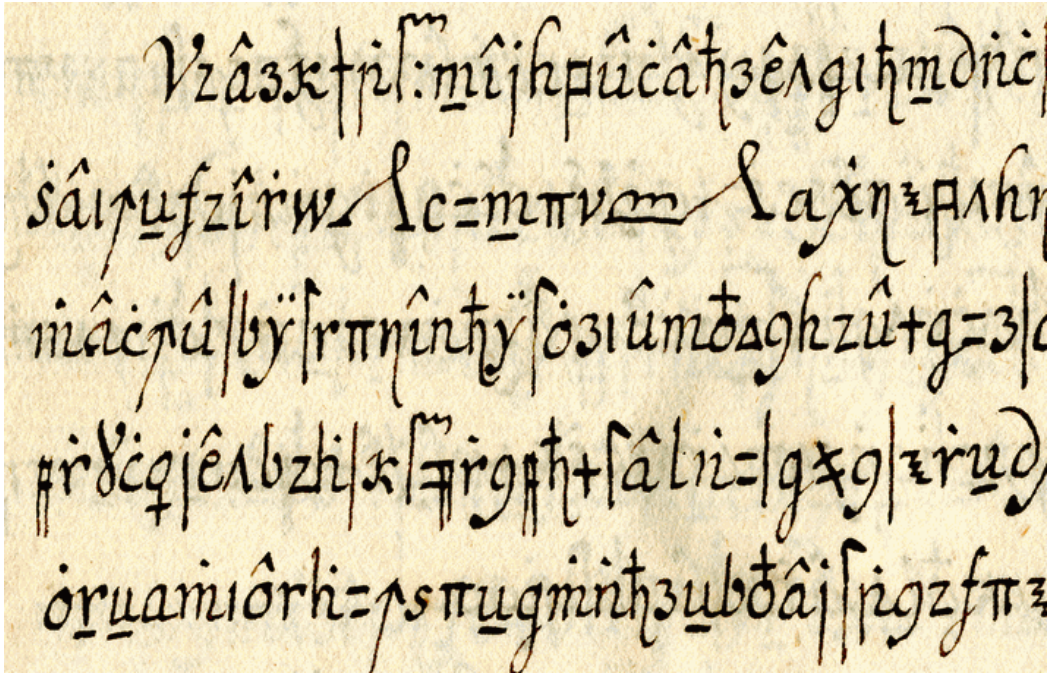
- An even distribution would look like the above
- But the single letter distribution of English is uneven

The letter distribution of English is uneven



- An even distribution would look like the above
- But the single letter distribution of Swedish is uneven

Copiale

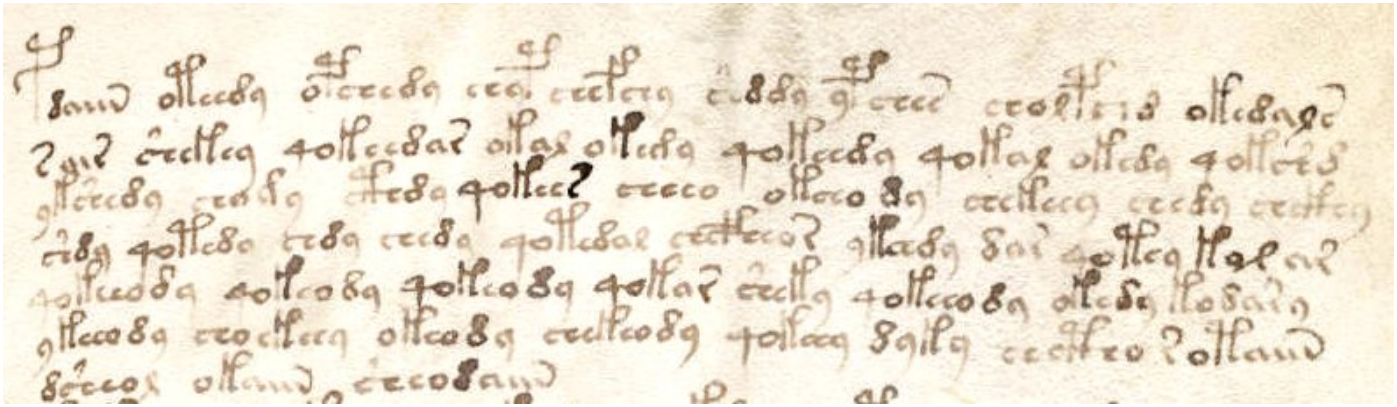


- Includes abstract symbols, Greek and Roman letters
- Broken in 2011, by researchers from USC and UU
- Breakthrough: unaccented Roman letters are whitespace
- Text is in German, from a secret society

The Voynich manuscript

- The Voynich manuscript is written on 15th century vellum
- $\sim$240 pages,
- Lots of illustrations, seems to be herbal, astronomical, maybe even pharmaceutical

The Voynich manuscript



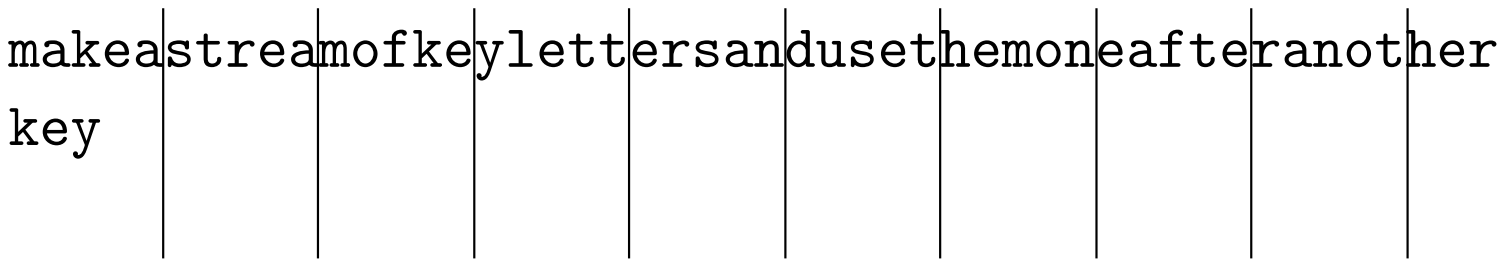
- An alphabet with 20–30 glyphs would account for almost all of the text
- There are some that occur only once or twice
- Statistical analysis of the text reveals patterns similar to those of natural languages.
- The text seems to be more repetitive than typical European languages.

Vigenere: change the substitution

- Principle used: Use Caesar letter by letter, but make sure that each plaintext letter maps to several different crypto letters, the actual letter depending on the position in the plaintext. Then, even methods for breaking simple substitution won't work
- Make a stream of key letters, and use them one after another for Caesar encryption of the corresponding plaintext letter.

Vigenere: change the substitution

- First: true Vigenère
- Make a stream of key letters, and use them one after another for Caesar encryption of the corresponding plaintext letter
- Start with a short key, easy to remember, and continue with the so far encrypted plaintext



Vigenere: change the substitution

- First: true Vigenère
- Make a stream of key letters, and use them one after another for Caesar encryption of the corresponding plaintext letter
- Start with a short key, easy to remember, and continue with the so far encrypted plaintext

m	a	k	e	a	s	t	r	e	a	m	o	f	k	e	y	l	e	t	t	e	r	s	a	n	d
k	e	y	m	a	k	e	a	s	t	r	e	a	m	o	f	k	e	y	l	e	t	t	e	r	s
w	e	i	q	a	c	x	r	w	t	d	s	f	w	s	d	v	i	r	e	i	k	l	e	e	v

u	s	e	t	h	e	m	o	n	e	a	f	t	e	r	a	n	o	t	h	e	r				
a	n	d	u	s	e	t	h	e	m	o	n	e	a	f	t	e	r	a	n	o	t				
f	h	n	z	i	f	v	r	q	o	s	x	e	w	t	r	f	t	u	s	k					

Vigenere: change the substitution

- “Normal” Vigenère
- Make a stream of key letters, and use them one after another for Caesar encryption of the corresponding plaintext letter
- Start with a short key, easy to remember, and repeat that over and over

m	a	k	e	a	s	t	r	e	a	m	o	f	k	e	y	l	e	t	t	e	r	s	a	n	d
k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e
w	e	i	o	e	q	d	v	c	k	q	m	p	o	c	i	p	c	d	x	c	b	w	y	x	h

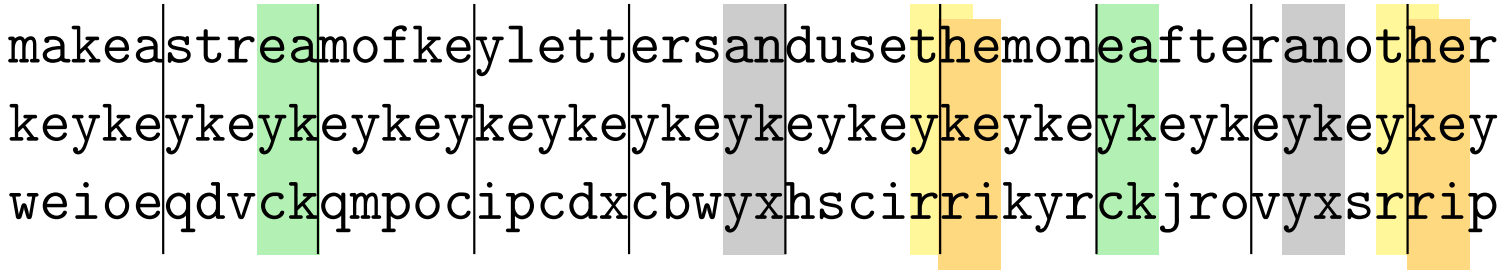
u	s	e	t	h	e	m	o	n	e	a	f	t	e	r	a	n	o	t	h	e	r				
y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y				
s	c	i	r	r	i	k	y	r	c	k	j	r	o	v	y	x	s	r	r	i	p				

Breaking Vigenère

```
makeastreamofkeylettersandusethemoneafteranother
keykeykeykeykeykeykeykeykeykeykeykeykeykeykeykey
weioeqdvckqmpocipcdxcbwyxhscirrikyrckjrovyxsrrip
```

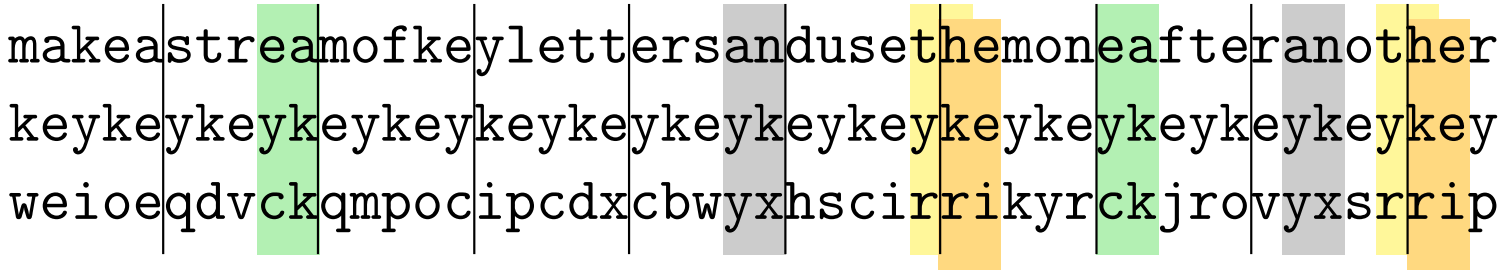
- Note that common (groups of) letters are likely to occur several times above the same position for the key.
- Trying a common syllable at all possible positions in the cryptogram will get several correct hits and thus expose a repeated portion of the key

Breaking Vigenère



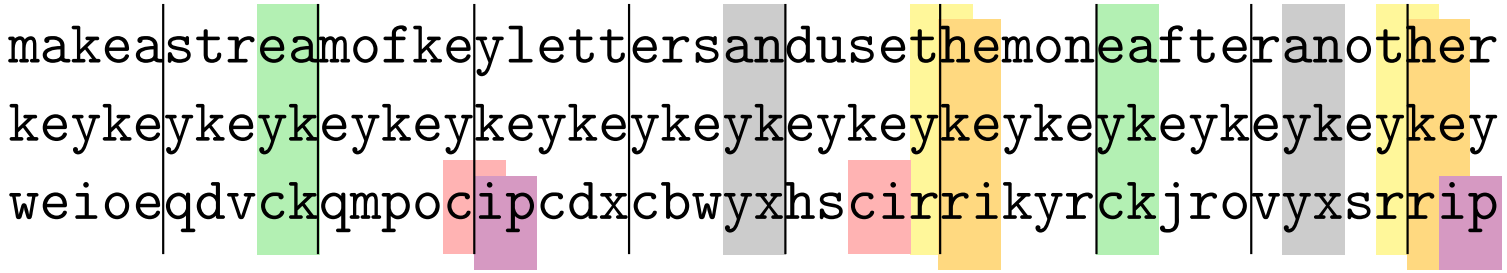
- The same pair of crypto letters (digrams) are often created by common plaintext hitting the same position in the key
- The distances between such pairs are different multiples of the key length
- Having the key length, you just treat all letters encrypted with the same key letter as a Caesar cipher

Breaking Vigenère



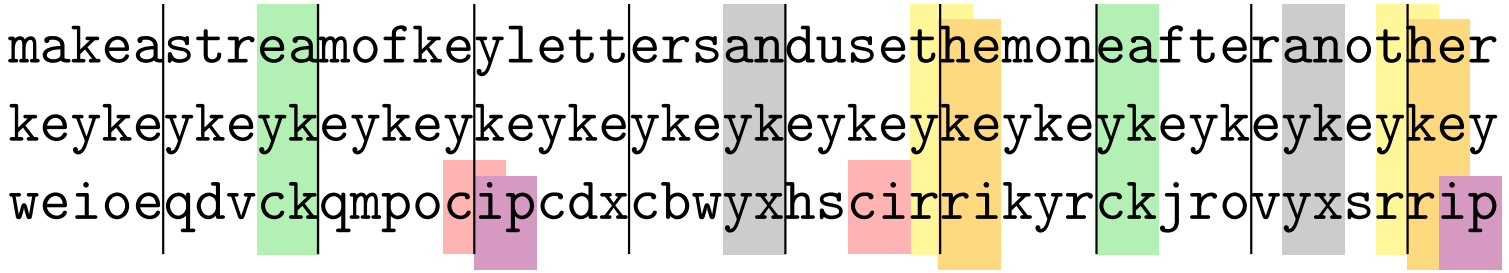
- In the cleartext, “ea” and “er” occur three times, and some other pairs occur twice
- “ea” has one repeated encryption, and also “th”, “he”, and “an”

Breaking Vigenère



- In the cleartext, “ea” and “er” occur three times, and some other pairs occur twice
- “ea” has one repeated encryption, and also “th”, “he”, and “an”
- (but “ci” and “ip” both decrypt to two different plaintexts)

Breaking Vigenère



- Find the distances between the repeated cryptoletter pairs, here 27, 13, 31, 18, 15, and 15
- Note that four out of six are divisible by 3
- Create three groups of crypto letters corresponding to each key letter, and identify the Caesar alphabet used for each by finding reasonable single letter statistics

The Playfair cipher

- Uses digrams (letter pairs) rather than single letters
- Create a 5x5 table of letters (“i” and “j” counts as the same)
- Fill it in, starting with a keyword (deleting repeated letters), and continue with the alphabet

p	l	a	y	f
i	r	b	c	d
e	g	h	k	m
n	o	q	s	t
u	v	w	x	z

The Playfair cipher

- Split your text into digrams, adding an “x” if a digram repeats a letter
- Three transformations are used for different cases:
- Same row: ciphertext is letters below
- Same column: letters to the right
- Otherwise: the other corners of the rectangle

p	l	a	y	f
i	r	b	c	d
e	g	h	k	m
n	o	q	s	t
u	v	w	x	z

larsson $\rightarrow$ la rs so nx $\rightarrow$ rb co xv su

The Playfair cipher

- This is a substitution cipher on two-letter blocks
- Cryptanalysis:
  - Use digram frequencies, e.g., both “re” and “er” are common in English, and they encrypt to a digram and its reverse
  - Each plaintext letter has only eight alternatives
  - Final part of the table is predictable

p	l	a	y	f
i	r	b	c	d
e	g	h	k	m
n	o	q	s	t
u	v	w	x	z

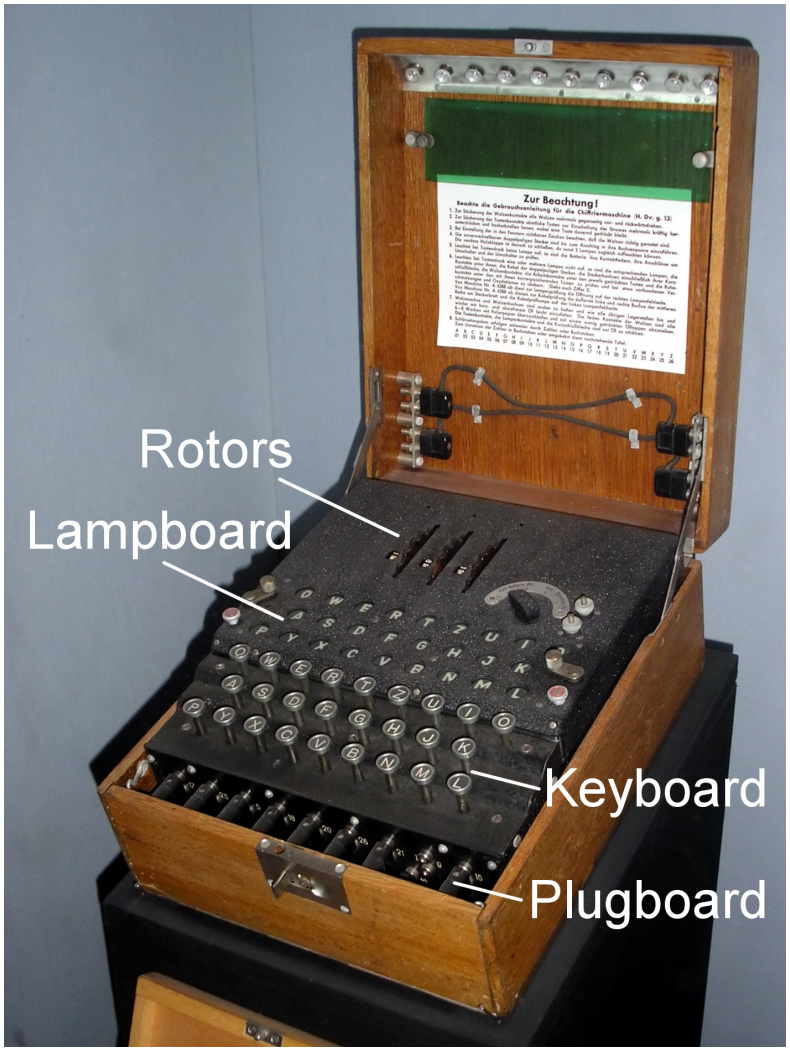
larsson $\rightarrow$ la rs so nx $\rightarrow$ rb co xv su

Is there a system to breaking codes and ciphers?

- Codes were regularly broken but never in a consistent way until the 19:th century. More like “learning the language” before that time.
- People were slowly becoming aware that letter (and digram) statistics are important.
- Examples include: ADFGX, Nihilist, basic block ciphers, Vigenere with long keys, Enigma,. . .

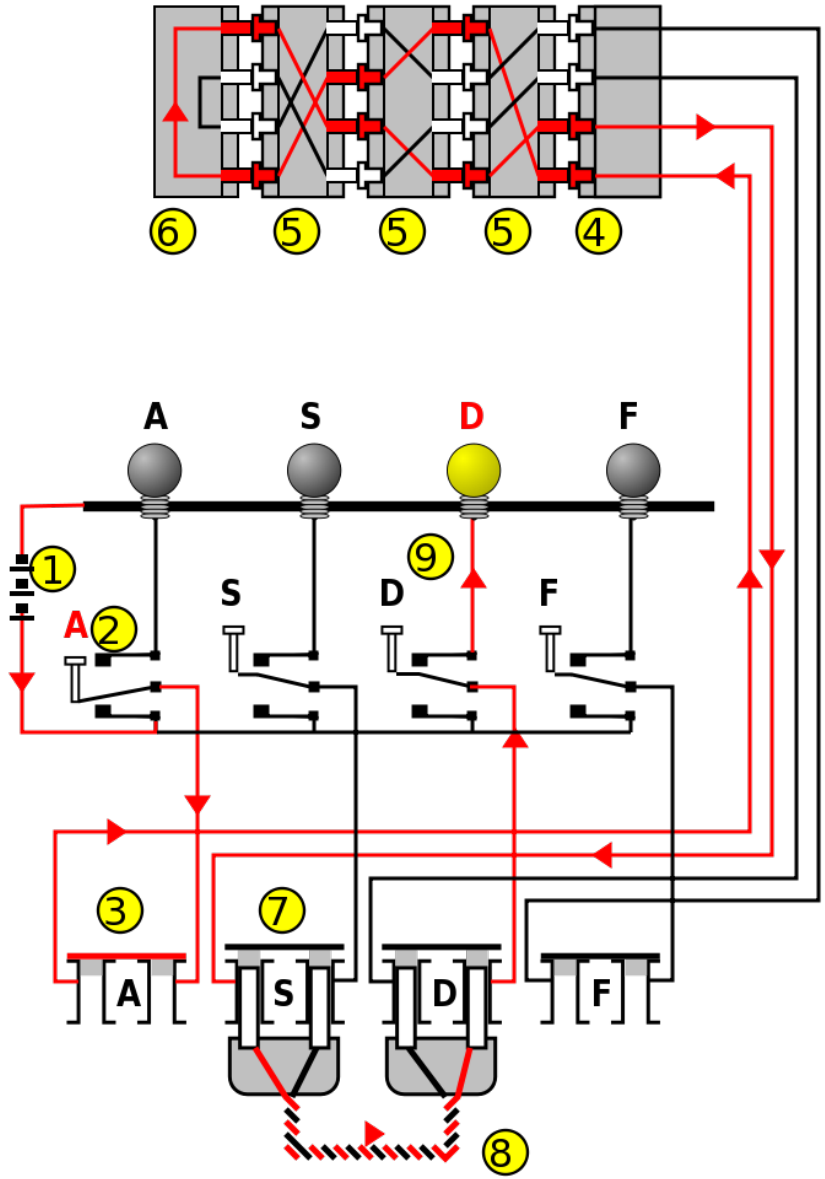
Enigma

- A “rotor machine”, constructed in the 1920s
- Broken in the 1930s by Polish cryptologists
- The techniques were passed to the British *only two months* before Germany invaded
- That it was broken was kept secret for almost 30 years after the war ended



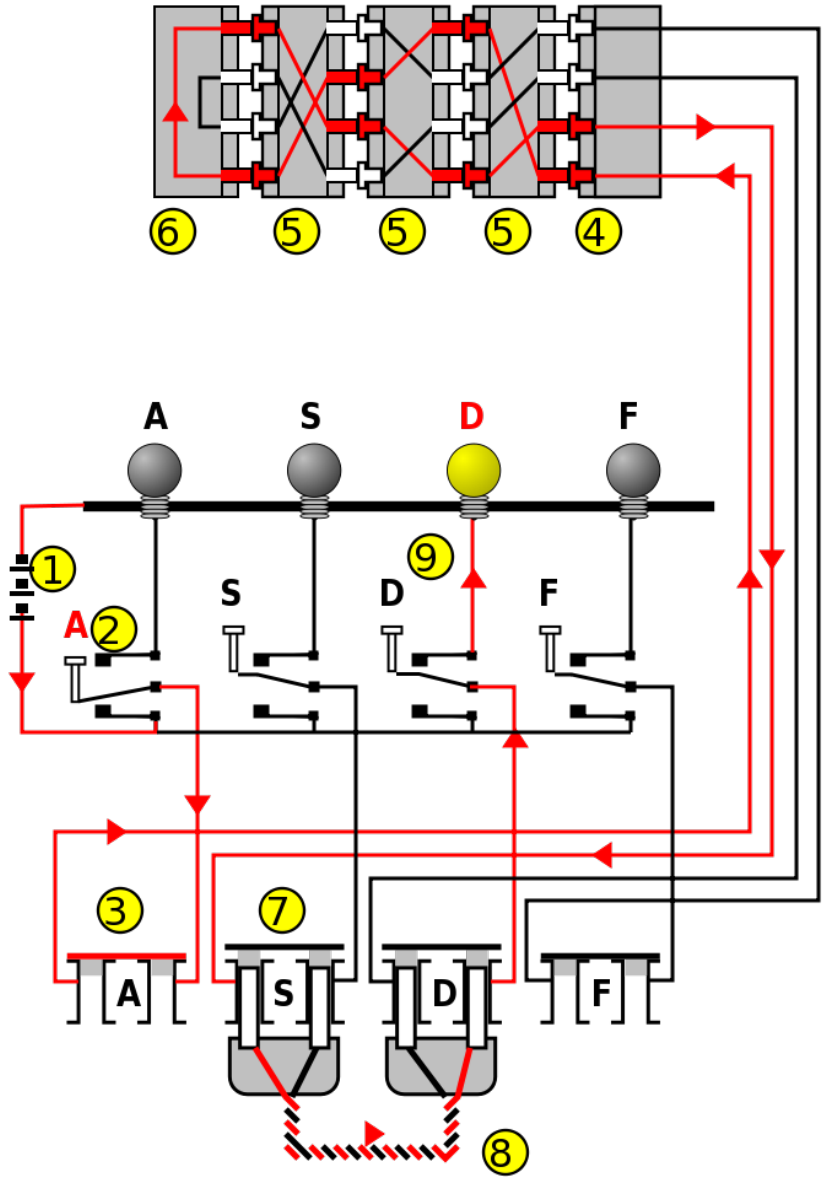
Enigma

- Same principle as Vigenere
- The key is
  - what rotors are used
  - their starting position
  - what letters are exchanged
- This is expanded to a long sequence of substitutions

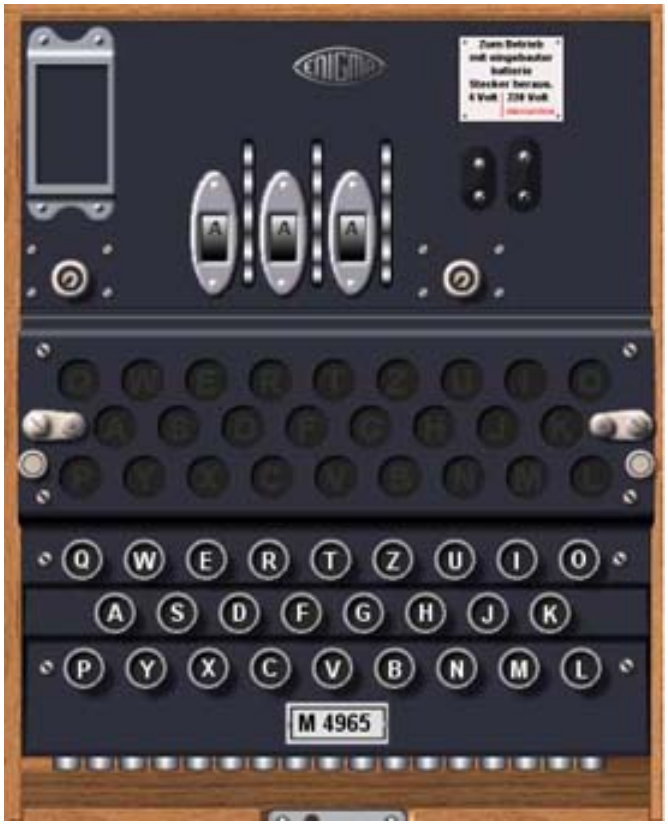


Enigma

- Three rotors: 6 combinations (later three-out-of-five: 60 combinations)
- Each rotor has 26 positions, and $26^3 = 17576$
- Total 105456 (or 1054560) combinations
- Plugboard has 100391791500 combinations

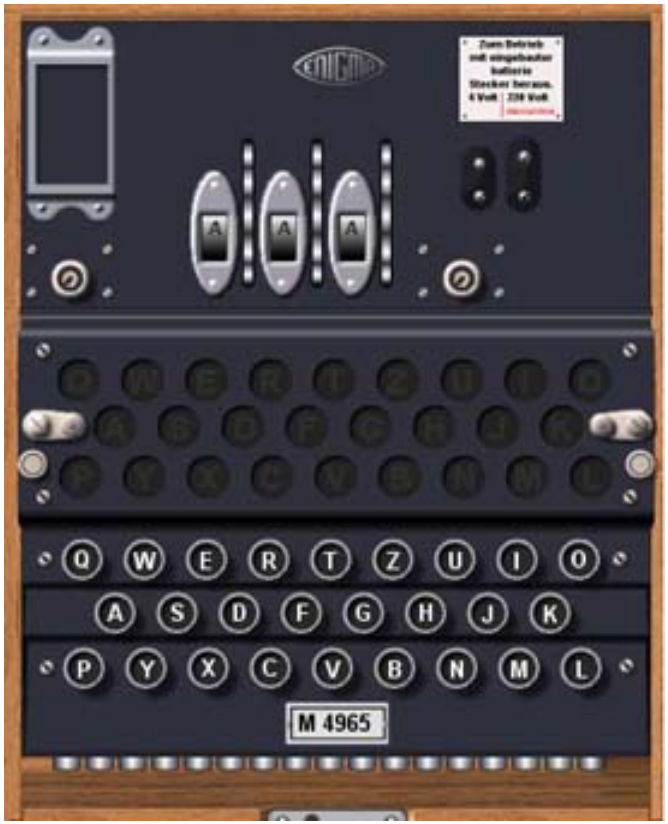


Breaking the Enigma



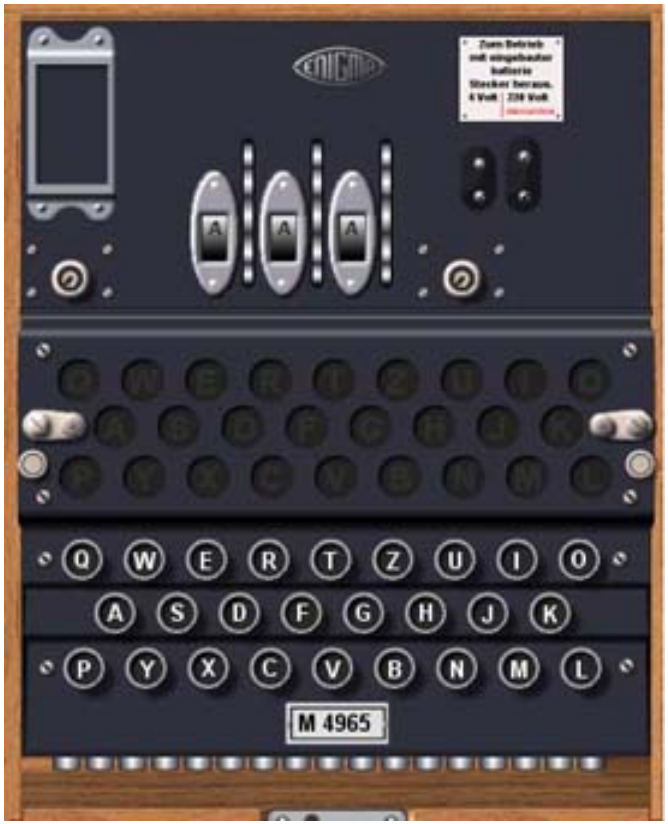
- Each Enigma operator was given a codebook with the daily settings for the next month
- If these had been used as given, many messages would have been sent with each day's setting
- The first letter of each message would have used the same substitution cipher
- The second (third, fourth, . . .) letter would have used another substitution cipher
- These could be broken by using letter statistics (of German)

Breaking the Enigma



- Each message starts with a three-letter message key, encrypted with the setting for the day
- The message key is repeated to avoid transmission errors
- Statistical analysis won't work
- But the repetition enables a different attack

Breaking the Enigma

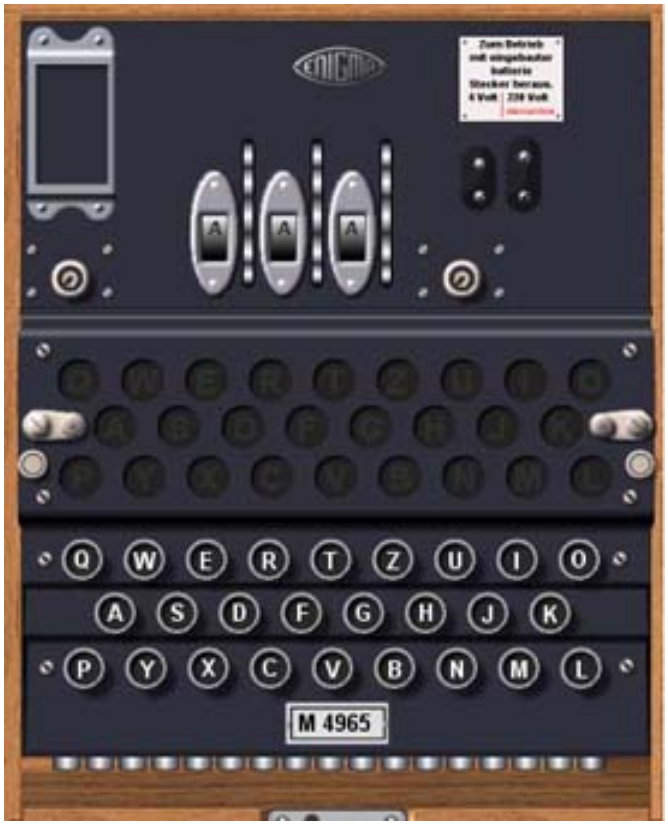


- Assume that you have received

  `dmqvbn`

- The first and fourth letters in each are the same letter, encrypted twice (with different substitution ciphers)
- The daily setting encrypts some unknown letter x to `d` in the first slot, and x to `v` in the fourth
- The Enigma is constructed such that if x goes to `d` (in the first slot), then `d` goes to x

Breaking the Enigma

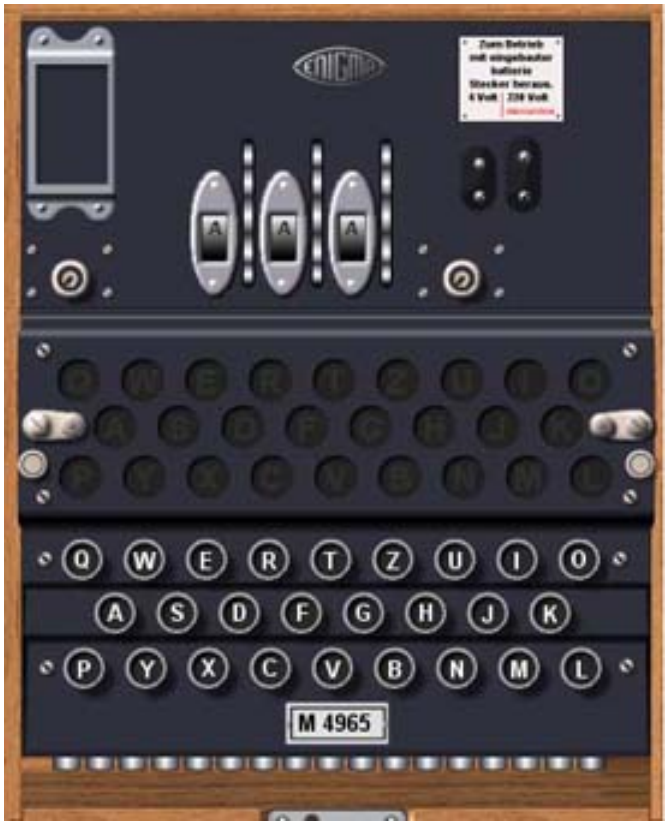


- Assume that you have received

  `dmqvbn`

- Encrypting `d` in the first slot gives the unknown x
- Encrypting x in the fourth slot gives `v`
- This eliminates the unknown x
- The first and fourth encryptions together maps `d` into `v`

Breaking the Enigma

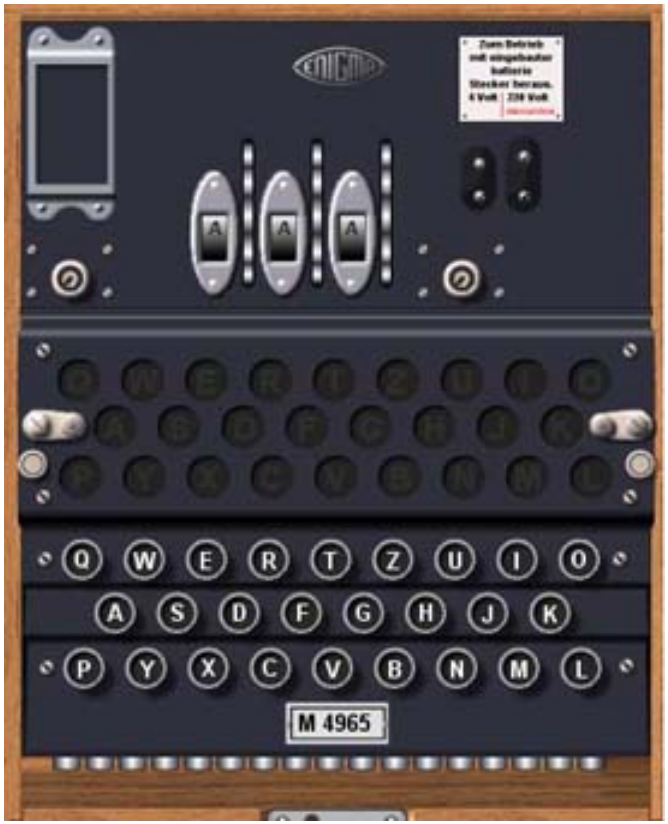


- Assume that you have received

```
dmqvbn
vonpuy
pucfmq
```

- The first and fourth encryptions together maps `d` into `v`
- The first and fourth encryptions together maps `v` into `p`
- The first and fourth encryptions together maps `p` into `f`

Breaking the Enigma

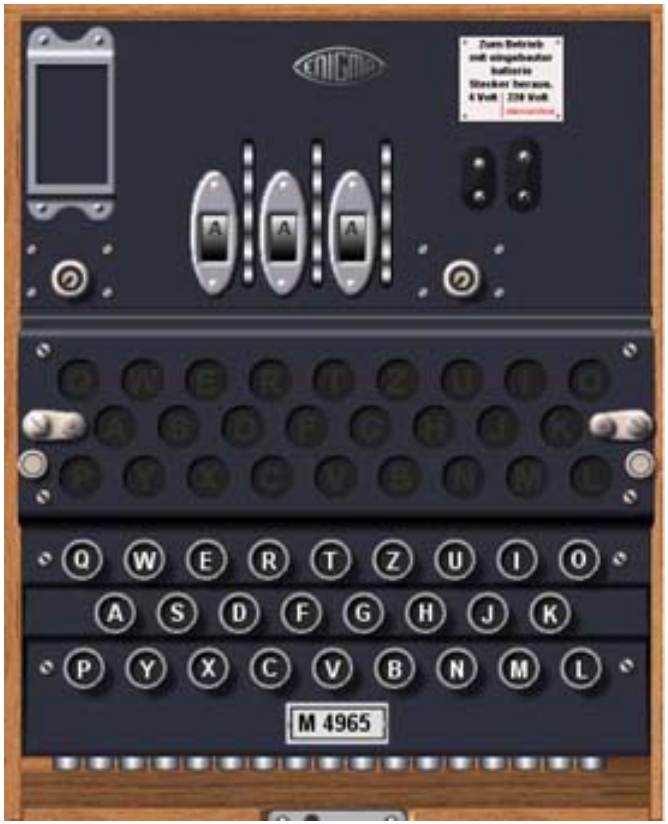


- Assume that you have received

```
dmqvbn
vonpuy
pucfmq
```

- The first and fourth encryptions together maps `d` into `v`
- The first and fourth encryptions together maps `v` into `p`
- The first and fourth encryptions together maps `p` into `f`
- With enough data, you'll find `d`$\rightarrow$`v`$\rightarrow$`p`$\rightarrow$`f`$\rightarrow$`k`$\rightarrow \ldots \rightarrow$`d`

Breaking the Enigma



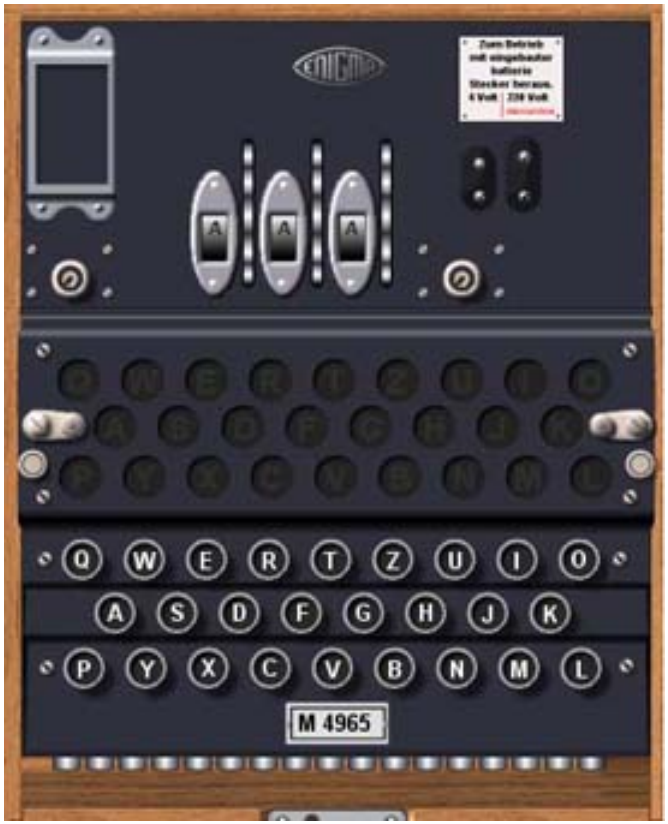
- Assume that you have received

```
dmqvbn
vonpuy
pucfmq
  ...
```

- The first and fourth encryptions together form a permutation

$$(dvpfkxgzyo)(eijmunqlht)(bc)(rw)(a)(s)$$

Breaking the Enigma



- The first and fourth encryptions together form a permutation

$$(dvpfkxgzyo)(eijmunqlht)(bc)(rw)(a)(s)$$

- The second and fifth encryptions together form a permutation

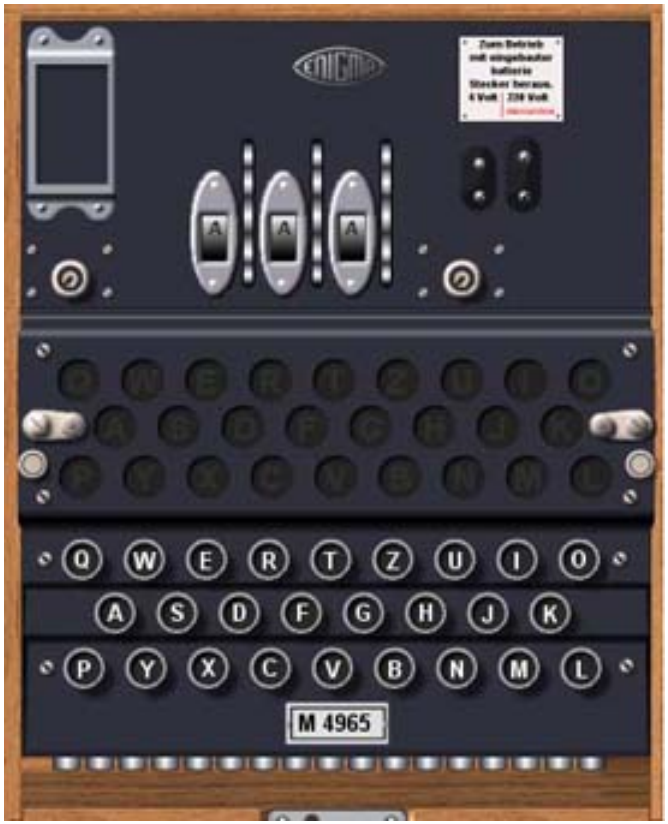
$$(blfqveoum)(hjpswizrn)(axt)(cgy)(d)(k)$$

- The third and sixth encryptions together form a permutation

$$(abviktjgfcqny)(duzrehlxwpsmo)$$

- These permutations are the same for every machine on a given day

Breaking the Enigma



- The first and fourth encryptions together form a permutation

$$(dvpfkxgzyo)(eijmunqlht)(bc)(rw)(a)(s)$$

- The plugboard interchanges some letters ($b \leftrightarrow p$, $f \leftrightarrow w$, ...)
- This changes the permutations
- But the lengths of the permutations stay the same: 10, 10, 2, 2, 1, 1

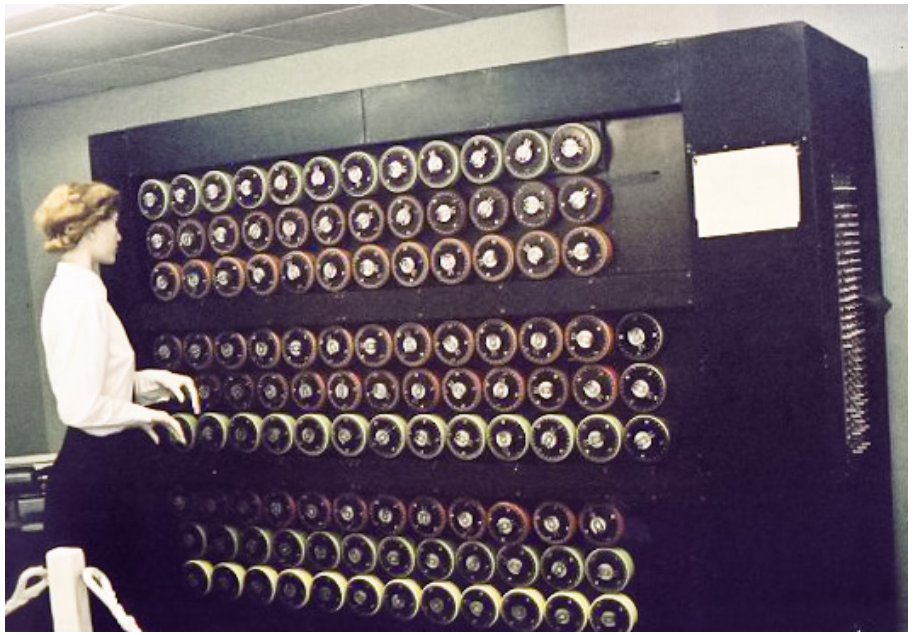
Breaking the Enigma

- The first and fourth encryptions together form a permutation

$$(dvpfkxgzyo)(eijmunqlht)(bc)(rw)(a)(s)$$

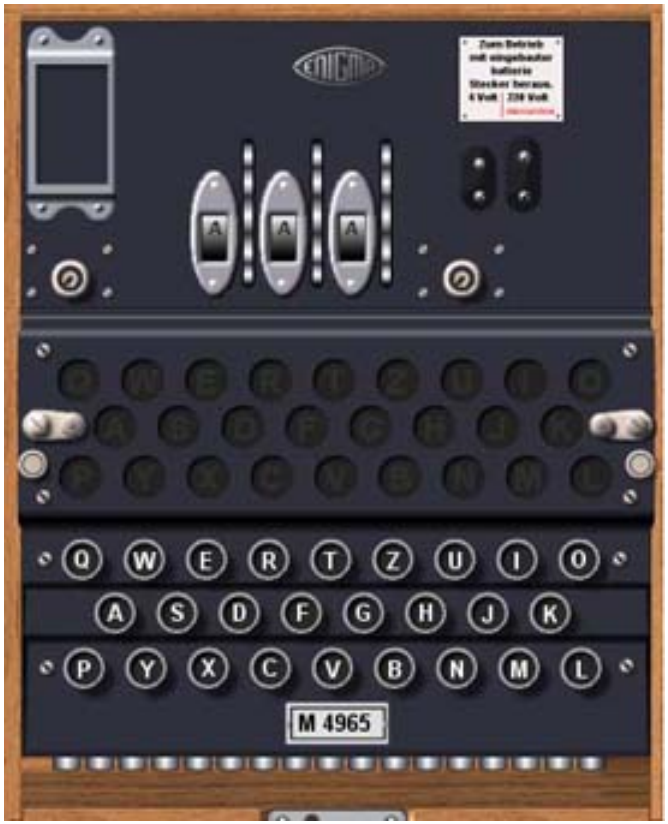
- These have lengths 10, 10, 2, 2, 1, 1
- The Polish cryptographers tabulated the permutation lengths for each daily key (105456 entries)
- Using this table, one only needed to check the subset with the correct permutation lengths
- . . . solving a substitution cipher for the plugboard

Breaking the Enigma



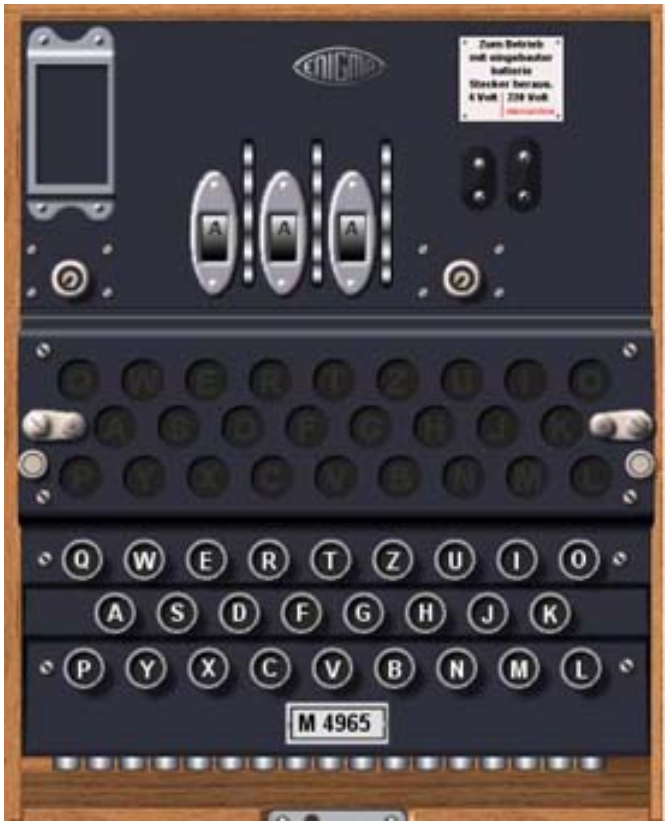
- During the war the search was made using Turing's "Bombes" at Bletchley park
- These were large electromechanical machines that needed lots of manual attention
- The output needed to be scanned for German (Italian, Japanese) words by hand

Key distribution for the Enigma



- The code book contained the daily settings:
  - Rotor order (later also: rotors used, three out of several)
  - Plugboard settings
  - Initial rotor positions
- Message key was encrypted twice with daily key
- Change September 1938: the initial rotor positions selected by operator and sent in the clear first
- Change May 1940: the message key only encrypted once

People are the problem



Even with the improved procedure there were many ways to crack Enigma

- Obscenities for keys
- Repeated parts of messages (“ANX” meaning “To:”)
- An operator sends a dummy message containing “LLLLL. . . ” which gave the English the wiring of the new rotor
- Few plugboard wires, rotors must not be in the same slot, . . .

Navajo code talkers

- Used in WWII by the US forces against the Japanese
- More of a code than an actual crypto
- Even sounds are unfamiliar to us, and hard to distinguish
- Hard to mimic, hard to send false messages
- The Japanese Imperial Army and Navy never cracked the spoken code

What we will do next lecture

- the One Time Pad (OTP), the only unbreakable cipher
- Shannon's information theory, that proves this
- the central role of statistics in cryptography

Register for lab 1 NOW