elicsir
Elevating Italian Computer Science Innovation and Research

UNIVERSITÀ DI TRENTO

Reed-Solomon: the hidden gem of QR Codes

Ludovico Cappellato

Mentor: Nicola Prezza

QR Code generalities

Full name: Quick Response Code

Year of birth: 1994

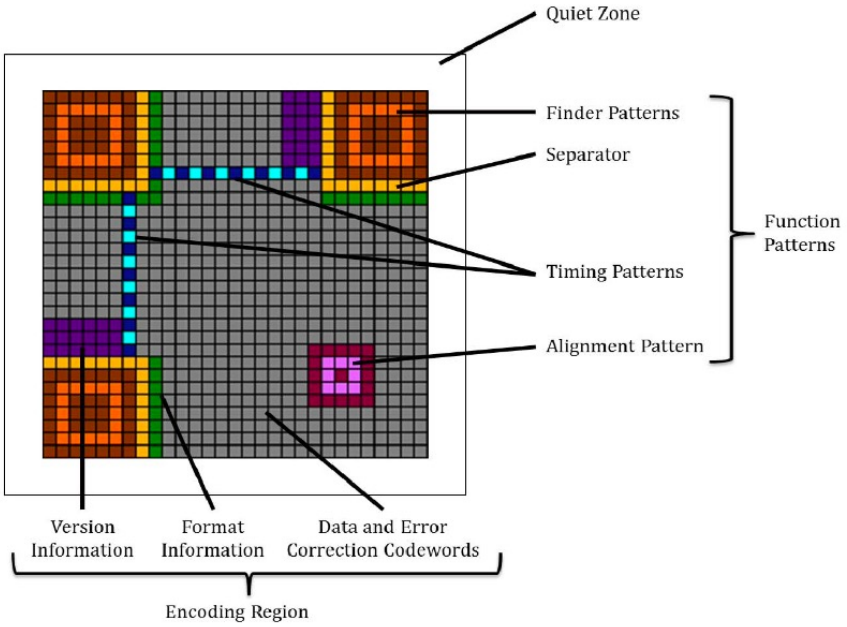
Place of birth: Japan

Favourite colours: Black & White

Size: from 21×21 to 177×177

Capacity: from 17 bytes to 3 KB

Main strength: avoids a lot of hand-copy of data



Not Only About Links

Smart QR Codes

There exists a plain-text standard to create more advanced functionalities

```
WIFI:S:<SSID>;
T:<WEP|WPA|nopass>;
P:<PASSWORD>;
H:<true|false|blank>;;
```

Wi-Fi access

```
PAGOPA|002|
<CODICE AVVISO>
|<CF ENTE
CREDITORE>|
<IMPORTO>
```

PagoPA

```
BEGIN:VEVENT
SUMMARY:<Meeting Title>
DTSTART:<20260501T090000Z>
DTEND:<20260501T100000Z>
LOCATION:<Text>
DESCRIPTION:<Text>
END:VEVENT
```

iCal plain text

```
tel:+<PREFIX>
<PHONE_NUMBER>
```

Phone Number

The Heart of QR Code

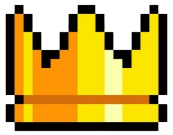
- Error correction
- Without it, almost useless
- Need to find a good error correction algorithm which balance strong error correction while preserving the compactness of QR Codes

Level L (Low): 7% recovery (Max capacity)

Level M (Medium): 15% recovery (Standard for general use)

Level Q (Quartile): 25% recovery

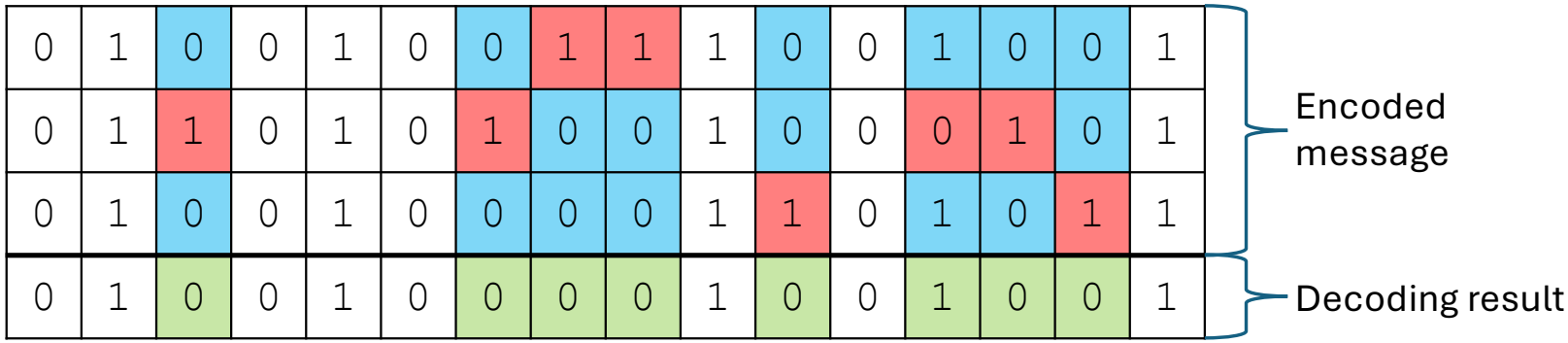
Level H (High): 30% recovery (Safest for industrial or outdoor use)



Plain Text → Message → Error correction redundancy → Codeword Placement

Message Repetitions

- What if we encode the message thrice and then take the average?



- Bad idea
  - Space: triples
  - Fail in message retrieval: $p(Fail \mid r) = 1 - \frac{\binom{n}{r} 3^r}{\binom{3n}{r}}$

Message Repetitions

- What if we encode the message thrice and then take the average?

0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1
0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1
0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1

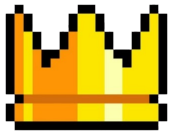
Encoded message

Decoding result

- Bad idea
  - Space: triples
  - Fail in message retrieval: $p(Fail \mid r) = 1 - \frac{\binom{n}{r} 3^r}{\binom{3n}{r}}$

The Heart of QR Code

- Reed-Solomon error correction
- Developed in 1960 under the anonymous name "Polynomial Codes Over Certain Finite Fields"
- Widely used in communication protocols, CDs, DVD, DVB, ...



Plain Text → Message → Error correction redundancy → Codeword Placement

Reed-Solomon error correction

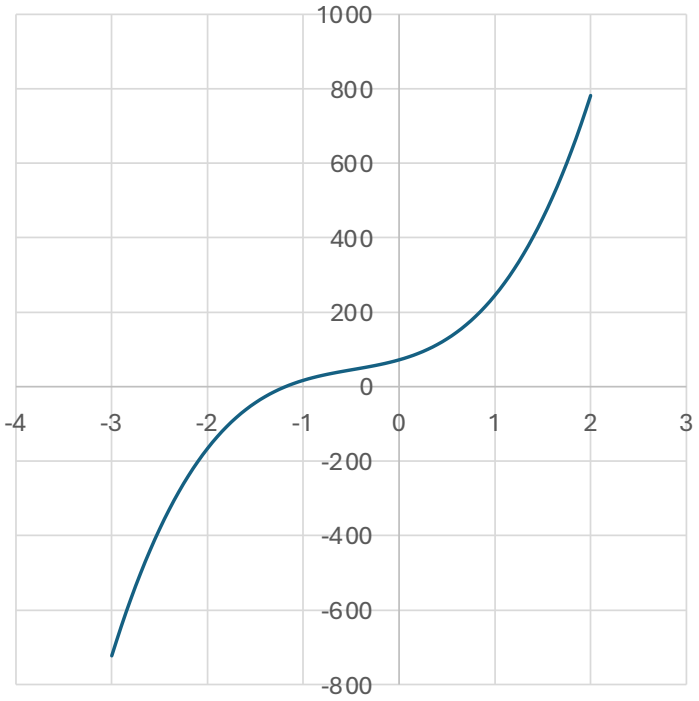
Reed-Solomon Intuition

- Step 0: We encode our message in a sequence of m symbols

  $\mathrm{m}=4 \quad (72,73,59,41)$

- Step 1: We use m to compose a (m-1)-degree polynomial using the sequence as its coefficients

$$(72, 73, 59, 41) \rightarrow P(x)$$
$$= 72 + 73x + 59x^2 + 41x^3$$



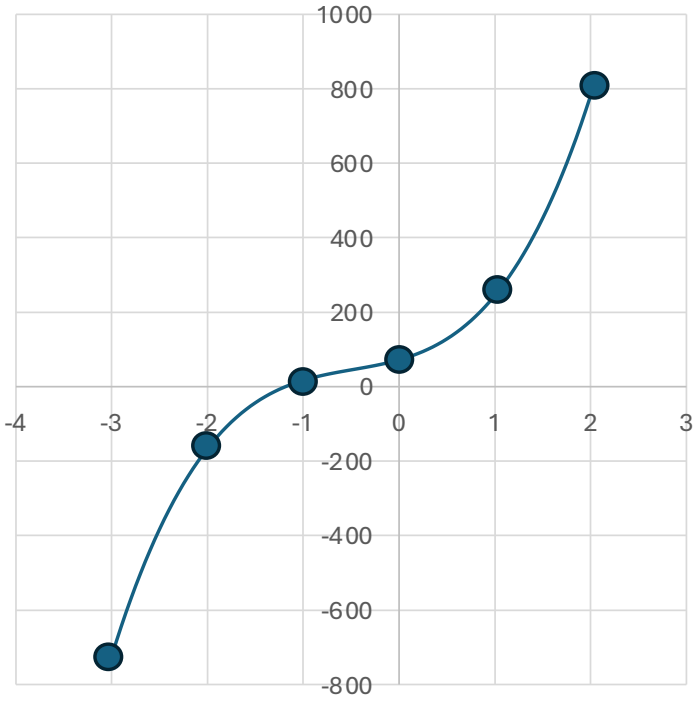
Reed-Solomon Intuition

- Step 2: We create a n-tuple of solutions to the polynomial, with n > m

$n=6$ $(-3,-2,-1,0,1,2)$

$(P(-3),P(-2),P(-1),P(0),$
$P(1),P(2)) \rightarrow$

$(-723,-166,17,72,245,782)$



- We use this n-sequence as our codeword

Reed-Solomon Intuition

- Step 3: We can reconstruct the original polynomial using polynomial interpolation algorithms
  - Since we need m points to uniquely identify a polynomial of grade m-1 we have $\binom{n}{m}$ possible determination of its coefficients, by solving independent systems of m linear equations

$n=6$ $(-3,-2,-1,0,1,2)$, $C = (-723,-166,17,72,245,782)$

$(-2,-166), (-1,17), (0,72), (1,245) \rightarrow (72,73,59,41)$

$(-2,-166), (-1,17), (0,72), (2,782) \rightarrow (72,73,59,41)$

...

Reed-Solomon Intuition

- Step 3: We can reconstruct the original polynomial using polynomial interpolation algorithms
  - If there are transmission errors we can still recover the original polynomial by looking at the largest number of determinations

$n=6$ (-3,-2,-1,0,1,2), $C = (-723,-166,17,72,245,782)$

$(-2,-166), (-1,17), (0,72), (1,245) \rightarrow (72,73,59,41)$

$(-2,-166), (-1,17), (0,72), (2,42) \rightarrow (\#\#,\#\#,\#\#,\#\#)$

$(-2,-166), (-1,17), (0,72), (-3,-723) \rightarrow (72,73,59,41)$

$(-2,-166), (-1,17), (1,245), (-3,-723) \rightarrow (72,73,59,41)$

...

Error Correction Power

Given m and n we have a total of $\binom{n}{m}$ determinations for the original m-tuple.
Assuming s errors we got $\binom{n-s}{m}$ correct determinations and
$\binom{n}{m} - \binom{n-s}{m}$ wrong determinations.

Lemma For s errors we can get at most $\binom{s+m-1}{m}$ determinations for a wrong m-tuple

Error Correction Power

Thus, by choosing the m-tuple with the higher number of determinations, we can recover the original tuple only when $\binom{n-s}{m} > \binom{s+m-1}{m}$

TL;DR: recover guaranteed if $s \leq \left\lfloor \frac{n-m}{2} \right\rfloor$

Comparison of Correction Power

- To sum up, with our dummy example m=4, n=6

Algorithm	1 byte error correction prob (correction rate)	2 bytes error correction prob (correction rate)	Transmitted data
Reed-Solomon	100% (17%)	0% (33%)	6 bytes
Message repetitions	52% (8%)	4% (17%)	12 bytes

- A bit unfair for Message repetitions

Comparison of Correction Power

- But in qrcodes we rarely find sparse bit filps
- If we have an error, it involves groups of continuous bits
- Reed-Solomon is more suited for error-burst correction

A Brief "Behind the Scenes"

- Actually, I oversimplified a bit the procedure, to make it more understandable in the whole picture
- Probably somewhere in the world some math students died due to my simplifications
- Let's try to put things in order

A Brief "Behind the Scenes"

- Reed-Solomon is defined over a Galois filed $GF(q)$ where q is the alphabet symbol size
- All the operations are defined inside the field
- Assuming q being a prime number, we can define all the operations in modulus q (in other words we are working in $\mathbb{Z}_p$ field)

ex. $\mathrm{GF}(5)$

$\mathrm{m}=3\ \ (2,1,4)\ \ P(x) = 2 + x + 4x^2\ \ \mathrm{n}=4\ \ (1,2,3,4)$

$\mathrm{P}(1)=7 \equiv 3 (\mathrm{mod}\ 5), \mathrm{P}(2)=20 \equiv 0 (\mathrm{mod}\ 5), \ldots$

$(2,1,4) \rightarrow (3,0,1,0)$

A Brief "Behind the Scenes"

- Bad news: 256 is not a prime number (and $\mathbb{Z}_{256}$ is not a field) 🙁
- ... but it is a prime power $q = p^n$ with p prime (2^8)
- we can still build a $GF(q) = GF(p)[X]/(P)$ by choosing an irreducible n-degree polynomial, typically $\mathrm{P} = x^8 + x^4 + x^3 + x + 1$.
- Our symbols will be mapped in $GF(2)$ and represented as a n-degree polynomial

ex. $73 = 01001001_2 \rightarrow x^6 + x^3 + 1$

A Brief "Behind the Scenes"

- Now we calculate our solutions 🙂

$P(x) = 72 + 73x + 59x^2 + 41x^3$

$P(2) = 72 \oplus (73 \otimes 2) \oplus (59 \otimes 4) \oplus (41 \otimes 8)$

$73 \otimes 2 = 01001001_2 \otimes 10_2 = (x^6 + x^3 + 1)(x) = x^7 + x^4 + x = 146$

$41 \otimes 8 = (x^5 + x^3 + 1)(x^3) = x^8 + x^6 + x^3 = x^4 + x^3 + x + 1 + x^6 + x^3$

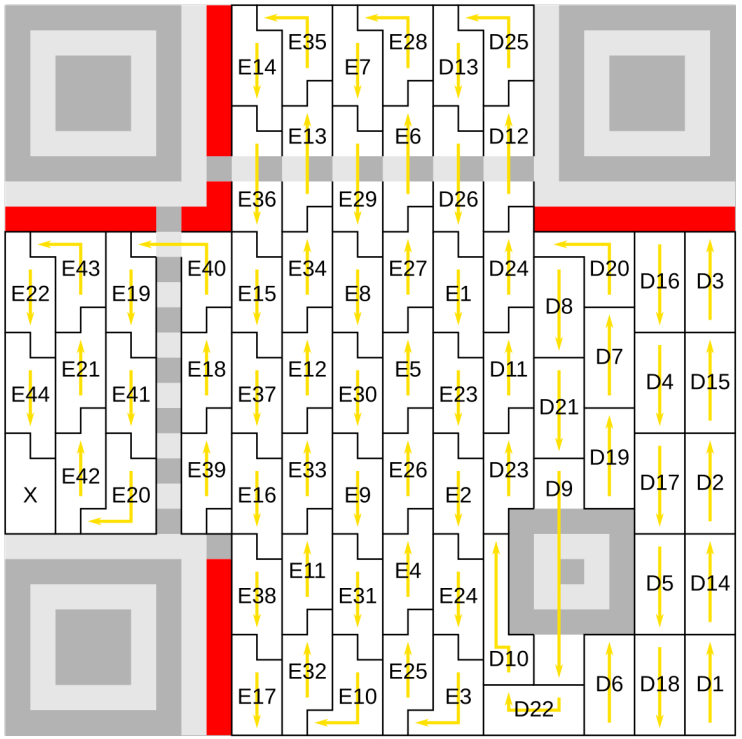
$= x^6 + x^4 + x + 1 = 83$

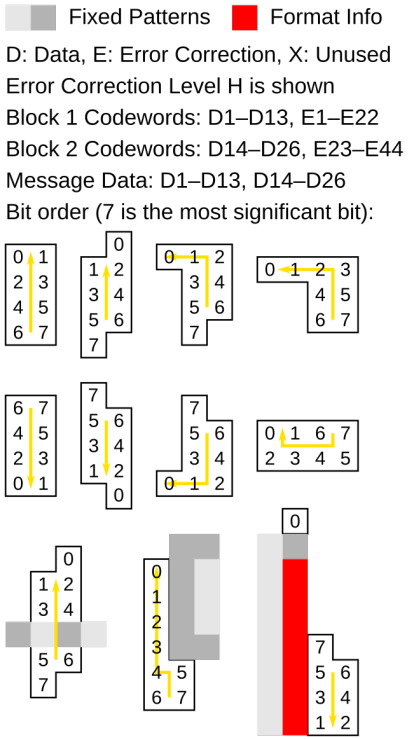
...

$P(2) = 101$

$n = 6$ $(0,1,2,3,4,5)$ $(P(0),P(1),P(2),P(3),P(4),P(5)) \rightarrow (72,19,101,200,68,29)$

Message Placement





No direct mapping between 0/1 and dark/white points, a mask is applied

Security Issues

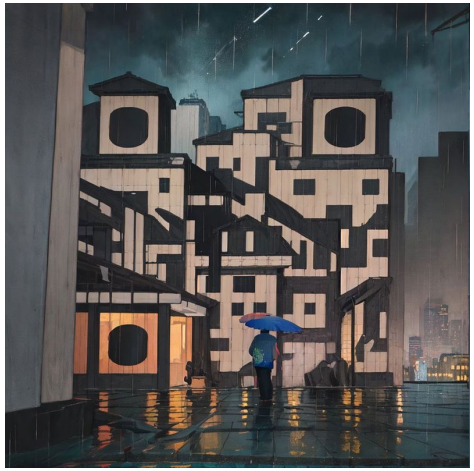
- Quishing
  - QR can automatically download software
  - Join a malicious Wi-Fi network
  - QR replacement attack (attagging)
  - Any thinkable social engineering attack...

- Third-party QR codes
  - Sensitive data disclosure

References

- Reed, Irving S., and Gustave Solomon. "Polynomial codes over certain finite fields." *Journal of the society for industrial and applied mathematics* 8.2 (1960): 300-304.
- Tiwari, Sumit. "An introduction to QR code technology." *2016 international conference on information technology (ICIT)*. IEEE, 2016.
- Artistic AI QRCodes https://qrbtf.com/en
- Open Source QRCode generator: https://gchq.github.io/CyberChef

Thank you for your attention



Proof of Lemma

For s errors we can get at most $\binom{s+m-1}{m}$ determinations for a wrong m-tuple

We look at a set of m elements from the n received values as the intersection of m hyperplanes.

Assuming a set with one wrong value we get one wrong determination. Hence, to have more determinations we would need at least another hyperplane intersecting at the same wrong point.

We can have at most s+m-1 hyperplanes intersecting at a single wrong point; therefore, we can have at most $\binom{s+m-1}{m}$ different sets determining the same wrong m-tuple. ■