## 6. Decoding Generalized Reed-Solomon Codes

## Decoding Generalized Reed-Solomon Codes

- We consider $\mathcal{C}_{\text {GRS }}$ over $\mathbb{F}_{q}$ with PCM

$$
H_{\mathrm{GRS}}=\left(\begin{array}{cccc}
1 & 1 & \ldots & 1 \\
\alpha_{1} & \alpha_{2} & \ldots & \alpha_{n} \\
\alpha_{1}^{2} & \alpha_{2}^{2} & \ldots & \alpha_{n}^{2} \\
\vdots & \vdots & \vdots & \vdots \\
\alpha_{1}^{\ell} & \alpha_{2}^{\ell} & \ldots & \alpha_{n}^{\ell} \\
\vdots & \vdots & \vdots & \vdots \\
\alpha_{1}^{r-1} & \alpha_{2}^{r-1} & \ldots & \alpha_{n}^{r-1}
\end{array}\right)\left(\begin{array}{cccc}
v_{1} & & & \\
& v_{2} & & 0 \\
0 & & \ddots & \\
& & & v_{n}
\end{array}\right)
$$

with $\alpha_{1}, \alpha_{2}, \ldots, \alpha_{n} \in \mathbb{F}_{q}^{*}$ distinct, and $v_{1}, v_{2}, \ldots, v_{n} \in \mathbb{F}_{q}^{*}$ (recall that $r=n-k=d-1$ ).

- Codeword $\mathbf{c}$ transmitted, word $\mathbf{y}$ received, with error vector

$$
\mathbf{e}=\left(e_{1} e_{2} \ldots e_{n}\right)=\mathbf{y}-\mathbf{c} .
$$

- $J=\left\{\kappa: e_{\kappa} \neq 0\right\}$ set of error locations.
- We describe an algorithm that correctly decodes y to $\mathbf{c}$, under the assumption $|J| \leq \frac{1}{2}(d-1)$.


## Syndrome Computation

- First step of the decoding algorithm: syndrome computation

$$
\begin{aligned}
\mathbf{S} & =\left(\begin{array}{c}
S_{0} \\
S_{1} \\
\vdots \\
S_{r-1}
\end{array}\right)=H_{\mathrm{GRS}} \mathbf{y}^{T}=H_{\mathrm{GRS}} \mathbf{e}^{T} \quad \begin{array}{c}
\ell \text { th row of } H_{\mathrm{GRS}}: \\
{\left[v_{1} \alpha_{1}^{\ell}, v_{2} \alpha_{2}^{\ell}, \ldots, v_{n} \alpha_{n}^{\ell}\right]}
\end{array} \\
S_{\ell} & =\sum_{j=1}^{n} y_{j} v_{j} \alpha_{j}^{\ell}=\sum_{j=1}^{n} e_{j} v_{j} \alpha_{j}^{\ell}=\sum_{j \in J} e_{j} v_{j} \alpha_{j}^{\ell}, \quad \ell=0,1, \ldots, r-1 .
\end{aligned}
$$

Example: For conventional RS codes, we have $\alpha_{j}=\alpha^{j-1}$ and $v_{j}=\alpha^{b(j-1)}$, so

$$
\begin{aligned}
S_{\ell}=\sum_{j=1}^{n} y_{j} \alpha^{(j-1)(b+\ell)}=y\left(\alpha^{b+\ell}\right), \quad \ell=0,1, \ldots, r-1 \\
\quad\left(\text { recall } \mathbf{c} \in \mathcal{C}_{\mathrm{RS}} \Leftrightarrow c\left(\alpha^{b+\ell}\right)=0, \ell=0,1, \ldots r-1\right)
\end{aligned}
$$

- Syndrome polynomial:

$$
S(x)=\sum_{\ell=0}^{r-1} S_{\ell} x^{\ell}=\sum_{\ell=0}^{r-1} x^{\ell} \sum_{j \in J} e_{j} v_{j} \alpha_{j}^{\ell}=\sum_{j \in J} e_{j} v_{j} \sum_{\ell=0}^{r-1}\left(\alpha_{j} x\right)^{\ell}
$$

## A Congruence for the Syndrome Polynomial

$$
S(x)=\sum_{j \in J} e_{j} v_{j} \sum_{\ell=0}^{r-1}\left(\alpha_{j} x\right)^{\ell}
$$

- We have

$$
\left(1-\alpha_{j} x\right) \sum_{\ell=0}^{r-1}\left(\alpha_{j} x\right)^{\ell}=1-\left(\alpha_{j} x\right)^{r} \equiv 1\left(\bmod x^{r}\right)
$$

Therefore, we can write

$$
\sum_{\ell=0}^{r-1}\left(\alpha_{j} x\right)^{\ell} \equiv \frac{1}{1-\alpha_{j} x}\left(\bmod x^{r}\right)
$$



$$
\begin{gathered}
S(x) \equiv \sum_{j \in J} \frac{e_{j} v_{j}}{1-\alpha_{j} x} \quad\left(\bmod x^{r}\right) \\
\left(\sum_{\text {empty }} \square \triangleq 0\right)
\end{gathered}
$$

## More Auxiliary Polynomials

- Error locator polynomial (ELP)

$$
\Lambda(x)=\prod_{j \in J}\left(1-\alpha_{j} x\right) \quad\left(\prod_{\text {empty }} \square \triangleq 1\right)
$$

- Error evaluator polynomial (EEP)

$$
\Gamma(x)=\sum_{j \in J} e_{j} v_{j} \prod_{m \in J \backslash\{j\}}\left(1-\alpha_{m} x\right)
$$

- $\Lambda\left(\alpha_{\kappa}^{-1}\right)=0 \quad \Longleftrightarrow \quad \kappa \in J \quad$ roots of EEP point to error locations
- $\Gamma\left(\alpha_{\kappa}^{-1}\right)=e_{\kappa} v_{\kappa} \prod_{m \in J \backslash\{\kappa\}}\left(1-\alpha_{m} \alpha_{\kappa}^{-1}\right) \neq 0$

$$
\Longrightarrow \quad \operatorname{gcd}(\Lambda(x), \Gamma(x))=1
$$

- The degrees of ELP and EEP satisfy

$$
\operatorname{deg} \Lambda=|J| \quad \text { and } \quad \operatorname{deg} \Gamma<|J|
$$

Of course, we don't know $\Lambda(x), \Gamma(x)$ : our goal is to find them

## Key Equation of GRS Decoding

Since $|J| \leq \frac{1}{2}(d-1)$, from $\operatorname{deg} \Lambda=|J|, \operatorname{deg} \Gamma<|J|$ we get
(1) $\operatorname{deg} \Lambda \leq \frac{1}{2}(d-1)$
and
(2) $\operatorname{deg} \Gamma<\frac{1}{2}(d-1)$

The ELP and the EEP are related by

$$
\begin{gathered}
\Gamma(x)=\sum_{j \in J} e_{j} v_{j} \prod_{m \in J \backslash\{j\}}\left(1-\alpha_{m} x\right)=\sum_{j \in J} e_{j} v_{j} \frac{\Lambda(x)}{1-\alpha_{j} x}=\Lambda\left(x \sum_{j \in J} \frac{e_{j} v_{j}}{1-\alpha_{j} x}\right) \\
\Longrightarrow(3) \quad \Lambda(x) S(x) \equiv \Gamma(x)\left(\bmod x^{d-1}\right)
\end{gathered}
$$

(1) $+(2)+(3)$ : key equation of GRS decoding

We have $S(x)$, and we know $d$. We want to solve for $\Lambda(x)$ and $\Gamma(x)$ satisfying (1) $+(2)+(3)$.

## Key Equation of GRS Decoding (cont.)

(1)
$\operatorname{deg} \Lambda \leq \frac{1}{2}(d-1)$
(2) $\operatorname{deg} \Gamma<\frac{1}{2}(d-1)$
(3)

$$
\Lambda(x) S(x) \equiv \Gamma(x) \quad\left(\bmod x^{d-1}\right)
$$

The coefficients of $\Lambda(x)$ and $\Gamma(x)$ solve the system of linear equations
$d-1\left(\begin{array}{ccccc}S_{0} & 0 & 0 & \cdots & 0 \\ S_{1} & S_{0} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ S_{\tau-1} & S_{\tau-2} & \cdots & S_{0} & 0 \\ \hline S_{\tau} & S_{\tau-1} & \cdots & S_{1} & S_{0} \\ S_{\tau+1} & S_{\tau} & \cdots & S_{2} & S_{1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ S_{d-2} & S_{d-3} & \cdots & S_{d-1-\tau} & S_{d-2-\tau}\end{array}\right)\left(\begin{array}{c}\lambda_{0} \\ \lambda_{1} \\ \lambda_{2} \\ \vdots \\ \lambda_{\tau}\end{array}\right)=\left(\begin{array}{c}\gamma_{0} \\ \gamma_{1} \\ \vdots \\ \gamma_{\tau-1} \\ 0 \\ 0 \\ \vdots \\ 0\end{array}\right)\left(\tau \triangleq\left\lfloor\frac{d-1}{2}\right\rfloor\right)$

- a set of $r=d-1$ linear equations in the coefficients of $\Lambda$ and $\Gamma$
- the last $\left\lfloor\frac{1}{2}(d-1)\right\rfloor$ equations depend only on $\Lambda$
- we can solve for $\Lambda$, find its root set $J$, then solve linear equations for $e_{j}$
- straightforward solution leads to $O\left(d^{3}\right)$ algorithm - we'll present an $O\left(d^{2}\right)$ one


## The Extended Euclidean Algorithm for polynomials

Given $a(x), b(x)$ over a field $\mathbb{F}$, with $a(x) \neq 0$ and $\operatorname{deg} a>\operatorname{deg} b$, the algorithm computes sequences of
remainders $r_{i}(x)$, quotients $q_{i}(x)$, and auxiliary polynomials $s_{i}(x), t_{i}(x)$

$$
\begin{aligned}
& r_{-1}(x) \leftarrow a(x) ; r_{0}(x) \leftarrow b(x) ; \\
& s_{-1}(x) \leftarrow 1 ; s_{0}(x) \leftarrow 0 ; \\
& t_{-1}(x) \leftarrow 0 ; t_{0}(x) \leftarrow 1 ; \\
& \text { for }\left(i \leftarrow 1 ; r_{i-1}(x) \neq 0 ; i++\right)\{ \\
& \quad q_{i}(x) \leftarrow r_{i-2}(x) \operatorname{div} r_{i-1}(x) ; \\
& \quad r_{i}(x) \leftarrow r_{i-2}(x)-q_{i}(x) r_{i-1}(x) ; \\
& \quad s_{i}(x) \leftarrow s_{i-2}(x)-q_{i}(x) s_{i-1}(x) ; \\
& \quad t_{i}(x) \leftarrow t_{i-2}(x)-q_{i}(x) t_{i-1}(x) ; \\
& \}
\end{aligned}
$$

- Let $\nu=$ largest $i$ such that $r_{i} \neq 0$. Then, $r_{\nu}(x)=\operatorname{gcd}(a(x), b(x))$.
- We also know that $s_{\nu}(x) a(x)+t_{\nu}(x) b(x)=\operatorname{gcd}(a(x), b(x))$ (often used to compute modular inverses).


## Properties of the Euclidean Algorithm Sequences

## Proposition (E1)

The following relations hold:

$$
\begin{aligned}
& \text { (i) For } i=-1,0, \ldots, \nu+1: \quad s_{i}(x) a(x)+t_{i}(x) b(x)=r_{i}(x) \\
& \text { (ii) For } i=0,1, \ldots, \nu+1: \quad \operatorname{deg} t_{i}+\operatorname{deg} r_{i-1}=\operatorname{deg} a
\end{aligned}
$$

Proof. By induction on $i . \square$

## Proposition (E2)

Suppose that $t(x), r(x) \in \mathbb{F}[x] \backslash\{0\}$ satisfy the following conditions:

$$
\begin{aligned}
& \text { (C1) } \operatorname{gcd}(t(x), r(x))=1 \\
& \text { (C2) } \operatorname{deg} t+\operatorname{deg} r<\operatorname{deg} a \\
& \text { (C3) } t(x) b(x) \equiv r(x)(\bmod a(x))
\end{aligned}
$$

Then, for some $h \in\{0,1, \ldots, \nu+1\}$ and a constant $c \in \mathbb{F}$, we have

$$
t(x)=c \cdot t_{h}(x) \quad \text { and } \quad r(x)=c \cdot r_{h}(x) .
$$

Proof. Standard polynomial manipulations, Proposition (E1), and recalling that the sequence $\operatorname{deg} r_{i}$ is strictly decreasing.

## Solving the Key Equation

- Apply the Euclidean algorithm with $a(x)=x^{d-1}$ and $b(x)=S(x)$.
- Let $\Lambda(x)$ and $\Gamma(x)$ play the roles of $t(x)$ and $r(x)$, respectively, in Proposition (E2). The definitions of $\Lambda$ and $\Gamma$, and the key equation, guarantee that conditions (C1)-(C3) are satisfied.
(C1) $\operatorname{gcd}(t(x), r(x))=\operatorname{gcd}(\Lambda(x), \Gamma(x))=1$
(C2) $\operatorname{deg} t+\operatorname{deg} r=\operatorname{deg} \Lambda+\operatorname{deg} \Gamma<\operatorname{deg} a=d-1$
(C3) $t(x) b(x) \equiv r(x) \bmod a(x) \Leftrightarrow \Lambda(x) S(x) \equiv \Gamma(x) \bmod x^{d-1}$
- By Proposition (E2), we have $\Lambda(x)=c \cdot t_{h}(x)$ and $\Gamma(x)=c \cdot r_{h}(x)$ for some index $h$ and scalar constant $c$.
How do we find index $h$ ?


## Theorem

The solution to the key equation is unique up to a scalar constant, and it is obtained with the Euclidean algorithm by stopping at the unique index $h$ such that

$$
\operatorname{deg} r_{h}<\frac{1}{2}(d-1) \leq \operatorname{deg} r_{h-1}
$$

Proof. Such an $h$ exists because $r_{i}$ is strictly decreasing. The degree properties follow from the propositions.

## Finding the Error Values

- Formal derivatives in finite fields: $\left[\sum_{i=0}^{s} a_{i} x^{i}\right]^{\prime}=\sum_{i=1}^{s} i a_{i} x^{i-1}$ $(a(x) b(x))^{\prime}=a^{\prime}(x) b(x)+a(x) b^{\prime}(x) \quad$ (not surprising)
- For the ELP, we have

$$
\Lambda(x)=\prod_{j \in J}\left(1-\alpha_{j} x\right) \quad \Longrightarrow \quad \Lambda^{\prime}(x)=\sum_{j \in J}\left(-\alpha_{j}\right) \prod_{m \in J \backslash\{j\}}\left(1-\alpha_{m} x\right)
$$

and, for $\kappa \in J$,

$$
\begin{aligned}
\Lambda^{\prime}\left(\alpha_{\kappa}^{-1}\right) & =-\alpha_{\kappa} \prod_{m \in J \backslash\{\kappa\}}\left(1-\alpha_{m} \alpha_{\kappa}^{-1}\right), \\
\Gamma\left(\alpha_{\kappa}^{-1}\right) & =e_{\kappa} v_{\kappa} \prod_{m \in J \backslash\{\kappa\}}\left(1-\alpha_{m} \alpha_{\kappa}^{-1}\right)
\end{aligned}
$$

- Therefore, for all error locations $\kappa \in J$, we obtain

$$
e_{\kappa}=-\frac{\alpha_{\kappa}}{v_{\kappa}} \cdot \frac{\Gamma\left(\alpha_{\kappa}^{-1}\right)}{\Lambda^{\prime}\left(\alpha_{\kappa}^{-1}\right)}
$$

Forney's algorithm for error values

## Summary of GRS Decoding

Input: received word $\left(y_{1} y_{2} \ldots y_{n}\right) \in \mathbb{F}_{q}^{n}$.
Output: error vector $\left(e_{1} e_{2} \ldots e_{n}\right) \in \mathbb{F}_{q}^{n}$.
(1) Syndrome computation: Compute the polynomial $S(x)=\sum_{\ell=0}^{d-2} S_{\ell} x^{\ell}$ by

$$
S_{\ell}=\sum_{j=1}^{n} y_{j} v_{j} \alpha_{j}^{\ell}, \quad \ell=0,1, \ldots, d-2
$$

(2) Solving the key equation: Apply Euclid's algorithm to $a(x) \leftarrow x^{d-1}$ and $b(x) \leftarrow S(x)$ to produce $\Lambda(x) \leftarrow t_{h}(x)$ and $\Gamma(x) \leftarrow r_{h}(x)$, where $h$ is the smallest index $i$ for which $\operatorname{deg} r_{i}<\frac{1}{2}(d-1)$.
(3) Forney's algorithm: Compute the error locations and values by

$$
e_{j}=\left\{\begin{array}{cl}
-\frac{\alpha_{j}}{v_{j}} \cdot \frac{\Gamma\left(\alpha_{j}^{-1}\right)}{\Lambda^{\prime}\left(\alpha_{j}^{-1}\right)} & \text { if } \Lambda\left(\alpha_{j}^{-1}\right)=0 \\
0 & \text { otherwise }
\end{array} \quad, \quad j=1,2, \ldots, n\right.
$$

Complexity: 1. $O(d n) \quad$ 2. $O((|J|+1) d) \quad$ 3. $O((|J|+1) n)$

## Schematic for GRS Decoder



## Finding Roots of the ELP (RS Codes)

Chien search for RS codes $\left(\alpha_{j}=\alpha^{j-1}, 1 \leq j \leq n\right)$


At clock cycle \#j, the cell labeled $\Lambda_{i}$ contains

$$
\Lambda_{i} \alpha^{-i(j-1)}, 0 \leq i \leq \tau
$$

and the output of the circuit is

$$
\begin{aligned}
& \sum_{i=0}^{\tau} \Lambda_{i} \alpha^{-i(j-1)} \\
= & \Lambda\left(\alpha^{-(j-1)}\right)=\Lambda\left(\alpha_{j}^{-1}\right), 1 \leq j \leq n
\end{aligned}
$$

## Other Decoding Algorithms

Many decoding algorithms and variants have been developed over the years. We mention a few of the most important ones.

- Berlekamp algorithm [1967] (also referred to as Berlekamp-Massey due to a clearer description and improvements by Massey [1969]): first efficient solution of the key equation, using Newton's identities and solving for shortest recurrence that generates the syndrome sequence. Complexity comparable to the Euclidean algorithm.
- Welch-Berlekamp [1986]: Solves key equation starting from remainder syndrome $y(x)(\bmod g(x))$, without computing power sums. Akin to continued fractions and Padé approximations.
- List decoding: Decodes beyond $\tau=\left\lfloor\frac{1}{2}(d-1)\right\rfloor$ errors, producing a list of candidate decoded codewords. Very often, the coset leader is unique even beyond $\tau$. Dates back to the '50s, but has gotten recent focus due to elegant and efficient algorithms by Sudan ['97], Guruswami-Sudan ['99] and others.
- Soft decoding: Information on the reliability of the symbols is available. Can lead to significant gains in decoding performance.


## Applications: PDF417 bar code

## $\square$

PREFERACCESS

FREQUENT FLYER

ORDER ID: ACY1A1
ETICKET: 2302182845474
SEQ: 92

| TERMINAL | GATE <br> $* * * *$ | GROUP <br> 2 | SEAT <br> 2 E | BOARDING BEGINS AT <br> $14: 43$ |
| :--- | :---: | :---: | :---: | :---: |

**gATE CLOSES 10 MINUTES PRIOR TO DEPARTURE// CIERRE DE PUERTA 10 HINUTOS ANTES DE LA SALIIAA DEL VUELO **


PDF417: A multi-row, 1D bar code (PDF: $\underline{\text { Portable } \underline{\text { Data }} \text { File). }}$

## PDF417 bar code structure



Codeword: an alternating pattern of 4 bars and 4 spaces, of varying widths, satisfying some constraints (e.g. width $\leq 6$ ). Total width: 17; 417 comes from $4+17$.

- Basic global parameters (height, width, ECC level, etc.) are encoded in the left and right row indicators. A form of repetition coding (one copy per row).
- Consecutive rows use different sets of bar/space patterns (codewords). Each set has 929 codewords; 3 disjoint sets are used cyclically.
- Number of rows: $3 \leq h \leq 90$. Number of codewords per row: $1 \leq w \leq 30$ (all rows have the same number of codewords).
- Total number of codewords (all rows): $n \leq 928$.
- Using fixed tables, each codeword is mapped to a number in $\{0,1, \ldots, 928\}$, and interpreted as an element of GF (929) (929 is prime).


## PDF417: Codeword mapping

Table H1. The Bar-Space Sequence Table. Cluster 0

| $\underline{\text { b }}$ bsbsbs | val | bsbsbsbs | val | bsbsbsbs | val | bsbsbsbs | val | bsbsbsbs | val | bsbsbsbs | val |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 31111136 | 0 | 41111144 | 1 | 51111152 | 2 | 31111235 | 3 | 41111243 | 4 | 51111251 | 5 |
| 21111326 | 6 | 31111334 | 7 | 21111425 | 8 | 11111516 | 9 | 21111524 | 10 | 11111615 | 11 |
| 21112136 | 12 | 31112144 | 13 | 41112152 | 14 | 21112235 | 15 | 31112243 | 16 | 41112251 | 17 |
| 11112326 | 18 | 21112334 | 19 | 11112425 | 20 | 11113136 | 21 | 21113144 | 22 | 31113152 | 23 |
| 11113235 | 24 | 21113243 | 25 | 31113251 | 26 | 11113334 | 27 | 21113342 | 28 | 11114144 | 29 |
| 21114152 | 30 | 11114243 | 31 | 21114251 | 32 | 11115152 | 33 | 51116111 | 34 | 31121135 | 35 |
| 41121143 | 36 | 51121151 | 37 | 21121226 | 38 | 31121234 | 39 | 41121242 | 40 | 21121325 | 41 |
| 31121333 | 42 | 11121416 | 43 | 21121424 | 44 | 31121432 | 45 | 11121515 | 46 | 21121523 | 47 |
| 11121614 | 48 | 21122135 | 49 | 31122143 | 50 | 41122151 | 51 | 11122226 | 52 | 21122234 | 53 |
| 31122242 | 54 | 11122325 | 55 | 21122333 | 56 | 31122341 | 57 | 11122424 | 58 | 21122432 | 59 |
| 11123135 | 60 | 21123143 | 61 | 31123151 | 62 | 11123234 | 63 | 21123242 | 64 | 11123333 | 65 |
| 21123341 | 66 | 11124143 | 67 | 21124151 | 68 | 11124242 | 69 | 11124341 | 70 | 21131126 | 71 |
| 31131134 | 72 | 41131142 | 73 | 21131225 | 74 | 31131233 | 75 | 41131241 | 76 | 11131316 | 77 |
| : | : | : | : | : | : | : | : | : | : | : | : |

## PDF417: Error correction

- An error correction level, $s, 0 \leq s \leq 8$, is defined.
- The sequence of codewords (all rows) is interpreted as a code block in a $[k+r, k, r+1]$ shortened Reed Solomon code over GF(929), where
- $k$ is the number of codewords used for actual data.
- Raw data is mapped to codewords using various efficient modes depending on whether the data is numeric, text, binary, or mixed.
- One bar code can encode more than 1100 raw bytes, 1800 ASCII characters, or 2700 decimal digits, depending on the mode.
- $r=2^{s+1}$, so $r \in\{2,4,8,16,32,64,128,256,512\}$.
- $k+r \leq 928$.
- 2 check digits are reserved for detection; the rest (if any) are used for erasure and full error correction.
- The generator polynomial of the RS code is

$$
g(x)=\prod_{i=1}^{r}\left(x-3^{i}\right)
$$

3 is primitive in GF (929).

## Application: QR codes



Version 1: $21 \times 21$


Version 3: $29 \times 29$


Version 10: $57 \times 57$


Version 40: $177 \times 177$

A truly 2D, highly versatile bar code (array referred to as a symbol)

## Application: QR codes

Widespread use

- Product or part tracking (original motivation)
- Web links
- Restaurant menus
- Tickets
- Document verification
- ... etc.

Robust ECC allows for data recovery under significant damage, and also for graphic art customization.


Fully recoverable symbols

## QR codes: Versions ( = Sizes)

| Version | Size | Capacity | Version | Size | Capacity | Version | Size | Capacity | Version | Size | Capacity |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\underline{\mathrm{M} 1}$ | 11 | $41 / 2$ | $\underline{8}$ | 49 | 242 | $\underline{19}$ | 93 | 991 | $\underline{30}$ | 137 | 2185 |
| $\underline{\mathrm{M} 2}$ | 13 | 10 | $\underline{9}$ | 53 | 292 | $\underline{20}$ | 97 | 1085 | $\underline{31}$ | 141 | 2323 |
| $\underline{\mathrm{M} 3}$ | 15 | $161 / 2$ | $\underline{10}$ | 57 | 346 | $\underline{21}$ | 101 | 1156 | $\underline{32}$ | 145 | 2465 |
| $\underline{\mathrm{M} 4}$ | 17 | 24 | $\underline{11}$ | 61 | 404 | $\underline{22}$ | 105 | 1258 | $\underline{33}$ | 149 | 2611 |
| $\underline{1}$ | 21 | 26 | $\underline{12}$ | 65 | 466 | $\underline{23}$ | 109 | 1364 | $\underline{34}$ | 153 | 2761 |
| $\underline{2}$ | 25 | 44 | $\underline{13}$ | 69 | 532 | $\underline{24}$ | 113 | 1474 | $\underline{35}$ | 157 | 2876 |
| $\underline{3}$ | 29 | 70 | $\underline{14}$ | 73 | 581 | $\underline{25}$ | 117 | 1588 | $\underline{36}$ | 161 | 3034 |
| $\underline{4}$ | 33 | 100 | $\underline{15}$ | 77 | 655 | $\underline{26}$ | 121 | 1706 | $\underline{37}$ | 165 | 3196 |
| $\underline{5}$ | 37 | 134 | $\underline{16}$ | 81 | 733 | $\underline{27}$ | 125 | 1828 | $\underline{38}$ | 169 | 3362 |
| $\underline{6}$ | 41 | 172 | $\underline{17}$ | 85 | 815 | $\underline{28}$ | 129 | 1921 | $\underline{39}$ | 173 | 3532 |
| $\underline{7}$ | 45 | 196 | $\underline{18}$ | 89 | 901 | $\underline{29}$ | 133 | 2051 | $\underline{40}$ | 177 | 3706 |

Capacity $=$ number of main data bytes (including ECC)

## QR codes: masking



- An XOR mask is applied by the encoder to the raw data to minimize undesirable features (large areas of the same color, etc.).
- Several masks are tried, and the resulting array is scored for bad features. Mask with the best score is chosen.
- The choice is encoded in the symbol.


## QR codes: structure

Version 1 symbol: $21 \times 21$
locator patterns (larger symbols have more)


Format areas (2 copies): 5 bits of information, encoded with a [15, 5, 7] binary BCH code (small code, exhaustive decoding possible). Format info (5 bits):

- 2 bits: error correction level (4 levels: L, M, Q, H).
- 3 bits: masking pattern.


## QR codes: structure



Larger symbols (Version 7: $45 \times 45$ and higher) also carry version information: 6 bits, encoded with a binary $[18,6,8]$ code.
The code is derived from the $[23,12,7]$ (perfect) Golay code by taking the even codewords ([23, 11, 8]) and shortening.
As with format information, two copies are written.

## QR codes: main data with error correction


message bytes

Data is encoded using shortened $R S$ codes over GF(256).

| ECC | $n, n-k$ for <br> Level | redundancy in <br> general case |
| :---: | :---: | :---: |
| L | $21 \times 21$ symbol | 7 |
| M | 26,10 | $\approx 14 \%$ |
| Q | 26,13 | $\approx 30 \%$ |
| H | 26,17 | $\approx 50 \%$ |

For larger symbols:

- Data is broken up into multiple RS blocks ( $41 \times 41$ and larger)
- RS block length is limited so that $n-k \leq 30$ (complexity)
- RS blocks are interleaved

Examples:

| vers. | array <br> size | ECC <br> level | message <br> bytes | num. blocks <br> $\times(n, n-k)$ | ECC <br> level | message <br> bytes | num. blocks <br> $\times(n, n-k)$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 10 | $57 \times 57$ | L | 274 | $2 \times(86,18)$ <br> $2 \times(87,18)$ | Q | 154 | $6 \times(43,24)$ |
|  |  |  |  | 2956 | $19 \times(148,30)$ <br> $6 \times(149,30)$ | Q | 1666 |
| 40 | $177 \times 177$ | L |  |  | $34 \times(54,30)$ |  |  |
| $34 \times(55,30)$ |  |  |  |  |  |  |  |

