Polarized Nested Constructions

Ilya Dumer

University of California, Riverside, USA

ACCT 2014

Motivation

Construct explicit families of good polarized codes (will be done for $R \to 1$).

No such codes are known to date for error channels.

Design good moderate-length (1000-4000 bits) polarized codes. Polar codes and LDPC-type codes perform close to channel capacity only on long blocks.

Outline

- Recursive tree-like structure of RM codes.
- Tree paths as channels. Recalculations of channel reliabilities.
- Decoding and polarization on tree structures.
- Incomplete paths with ML decoding on end nodes.
- Design of nested polarized codes.

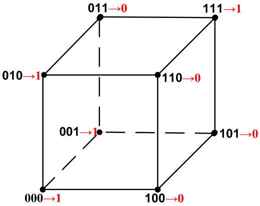
Reed-Muller (RM) codes $\mathcal{R}(r, m)$ of order $r = 1, ..., m$.

- Messages: polynomials $f^{(r)}(x_1, \ldots, x_m)$ of deg $\leq r$ in m Boolean variables.
- Codewords: outputs $\mathbb{F}_2^m \to \mathbb{F}_2$ of polynomials f
- Message $f^{(2)}(x_1, x_2, x_3) = x_2 x_3 + x_1 + 1$. Codeword: (11100001)

Length $n = 2^m$.

Dimension $k = \sum_{i=0}^{r} \binom{m}{i}$.

Minimum distance $d = 2^{m-r}$.

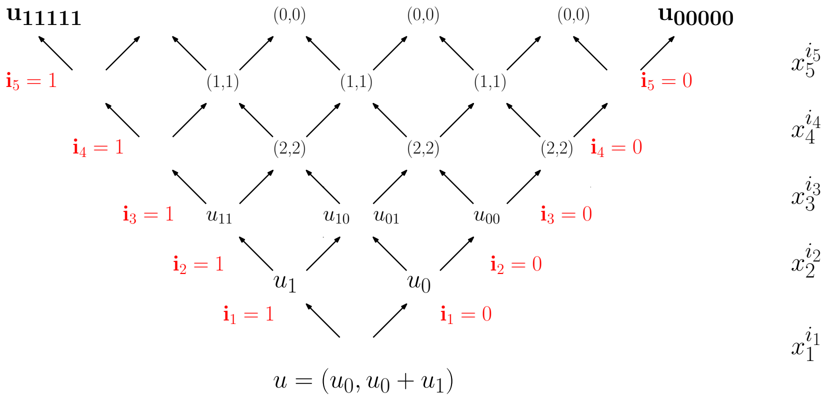


Partition: $f(x_1, ..., x_m) = f_0(x_2, ..., x_m) \;+\; x_1 f_1(x_2, ..., x_m)$

$= f_{00}(x_3, ..., x_m) \;+\; x_2\, f_{01}(x_3, ...) \;+\; x_1 f_{10}(x_3, ...) \;+\; x_1 x_2\, f_{11}(x_3, ...)$

$= \ldots = \Sigma_{i_1,\ldots,i_m}\; u_{i_1,\ldots,i_m}\; x_1^{i_1}\; \ldots\; x_m^{i_m}$

Full partition of (5,5) RM code into 32 paths



$x_1^{i_1}\ x_2^{i_2}\ ...\ x_5^{i_5}$ defines path $\psi = i_1 i_2 ... i_5$ from u to $u_{i_1 i_2 ... i_5}$

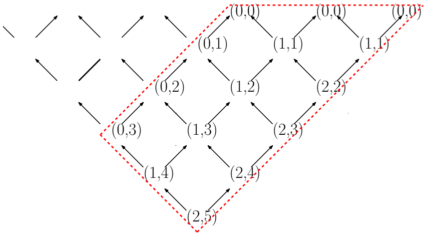
and row of gen. matrix with Hamming weight $2^{m-wt(\psi)}$

Plotkin $(\mathbf{u}, \mathbf{u}+\mathbf{v})$ construction for RM codes of order $r < m$

$$\underbrace{f^{(r)}(x_1,\ldots,x_m)}_{\mathbf{c}\in\mathcal{R}(r,m)} = \underbrace{f_0^{(r)}(x_1,\ldots,x_{m-1})}_{\mathbf{u}\in\mathcal{R}(r,m-1)} + \underbrace{x_1 f_1^{(r-1)}(x_2,\ldots,x_m)}_{\mathbf{v}\in\mathcal{R}(r-1,m-1)}$$

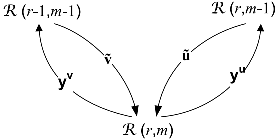
Generator matrix $G_r^m = \left[\begin{array}{c|c} G_r^{m-1} & G_r^{m-1} \\ \hline 0 & G_{r-1}^{m-1} \end{array}\right]$

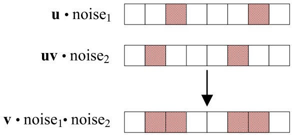
Codes $\mathcal{R}\,(r,m)$ map monomials f of $\deg(f) \le r$ and yield code weights $2^{m-\deg(f)}$ on the paths $(i_1,...,i_m)$ of weight $= \deg(f)$. Any vector $(\mathbf{u}, \mathbf{u}+\mathbf{v})$ has weight $\ge \min\{2wt(\mathbf{u}), wt(\mathbf{v})\}$. Codewords of min weight 2^{m-r} can generate $\mathcal{R}(r,m)$.



Decoding. Below we couple indices $i \in [1, n/2]$ and $(i) = i + n/2$, and use input alphabet $\{\pm 1\}$. Then $\mathbf{c} = (\underset{R(r,m-1)}{\mathbf{u}_i}, \underset{R(r-1,m-1)}{\mathbf{uv}_{(i)}})$

Received $(\mathbf{y}_i, \mathbf{y}_{(i)}) = (\mathbf{u} \cdot \mathbf{n}_i, \quad \mathbf{uv} \cdot \mathbf{n}_{(i)})$ with noise vectors $\mathbf{n}_i, \mathbf{n}_{(i)}$.
We will estimate $\mathbf{v} = \mathbf{y}_i \mathbf{y}_{(i)} \cdot \mathbf{n}_i \mathbf{n}_{(i)}$ and $\mathbf{u} = \mathbf{y}_i \mathbf{n}_i = \mathbf{v} \cdot \mathbf{y}_{(i)} \mathbf{n}_{(i)}$.





1. Form estimate $\mathbf{y}^v = \mathbf{y}_i \mathbf{y}_{(i)}$ of $\mathbf{v}$ and decode $\mathbf{y}^{\mathbf{v}} \Rightarrow \widetilde{\mathbf{v}}$.

   Here $\mathbf{y}_i \mathbf{y}_{(i)}$ has $\max$ of t errors with the mean $t(1 - t/n)$.

2. Form estimate $\mathbf{y}^u$ of $\mathbf{u}$ from $\mathbf{y}_i$ and $\widetilde{\mathbf{v}} \mathbf{y}_{(i)}$. Decode $\mathbf{y}^{\mathbf{u}} \Rightarrow \widetilde{\mathbf{u}}$.

- Recursion: Proceed to RM codes of order $r - 2, ..., 1$ (or $r = 0$).

Path recalculations for the received vector $\mathbf{y} = \left(\underset{\mathbf{u}}{\mathbf{c}_i \cdot \mathbf{n}_i}, \ \underset{\mathbf{uv}}{\mathbf{c}_{(i)} \cdot \mathbf{n}_{(i)}} \right)$

Consider posterior prob. of symbols c_i, $c_{(i)}$ and v_i, u_i

$$p_i \triangleq \Pr\{c_i = 1 \mid y_i\}, \qquad p_{(i)} \triangleq \Pr\{c = 1 \mid y_{(i)}\}$$
$$p_i^{\mathbf{v}} \triangleq \Pr\{v_i = 1 \mid y_i, \ y_{(i)}\}, \quad p_i^{\mathbf{u}} \triangleq \Pr\{u_i = 1 \mid y_i, \ y_{(i)}, \ v_i\}$$

Then $p_i^{\mathbf{v}}$ and $p_i^{\mathbf{u}}$ have a compact form via **probability offsets**:

$$g_i = 1 - 2p_i, \quad g_{(i)} = 1 - 2p_{(i)}, \qquad \widetilde{g}_{(i)} \triangleq g_{(i)}\tilde{v}_i$$
$$g_i^{\mathbf{v}} = g_i g_{(i)}, \qquad g_i^{\mathbf{u}} = \left(g_i + \widetilde{g}_{(i)}\right) / \left(1 + g_i \widetilde{g}_{(i)}\right)$$

The (product) $\mathbf{v}$-channel degrades $\mathbf{y}$-channel and the (repetition) $\mathbf{u}$-channel improves it. Specific changes depend on a $\mathbf{y}$-channel.

Channel	$\mathbf{y}$	$\mathbf{v}$	$\mathbf{u}$	Results
High noise*	$g \ll 1$	g^2	$\simeq 2g$	Big penalty for $\mathbf{v}$-path
Low noise	$p \ll 1$	$\simeq 2p$	$\simeq p^2$	Big gain for $\mathbf{u}$-path

*To reliably correct noise with offset g, a repetition code needs length $\succeq g^{-2}$.

BER of different information bits (channels) in recursive decoding

ALGORITHM $\mathcal{D}_r^m(\mathbf{g}_i, \mathbf{g}_{(i)})$

1. If $r = 1$, decode $\mathcal{R}(1, m)$. Else

2. Recursively decode $\mathcal{R}(r, m)$

Take $\mathbf{g}^{\mathbf{v}} = \mathbf{g}_i \mathbf{g}_{(i)}$; $\quad \widetilde{\mathbf{v}} = \mathcal{D}_{r-1}^{m-1}(\mathbf{g}^{\mathbf{v}})$.

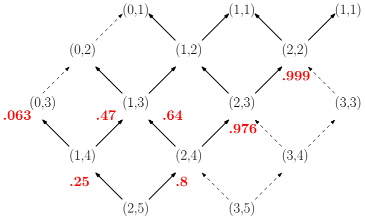
Take $\mathbf{g}^{\mathbf{u}} = \dfrac{\mathbf{g}_i + \widetilde{\mathbf{g}}_{(i)}}{1 + \mathbf{g}_i \widetilde{\mathbf{g}}_{(i)}}$; $\quad \widetilde{\mathbf{u}} = \mathcal{D}_r^{m-1}(\mathbf{g}^{\mathbf{u}})$.

Output $\widetilde{\mathbf{c}} = (\widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}\widetilde{\mathbf{v}})$.

$g_i^y = 0.5 \qquad g_i^{\mathbf{v}} = 0.25$ – degrading.

If $\widetilde{\mathbf{v}} = +1$: $\quad g_i^{\mathbf{u}} = 0.80$ – improving.

If $\widetilde{\mathbf{v}} = -1$: $\quad g_i^{\mathbf{u}} = 0 \quad$: – erasure

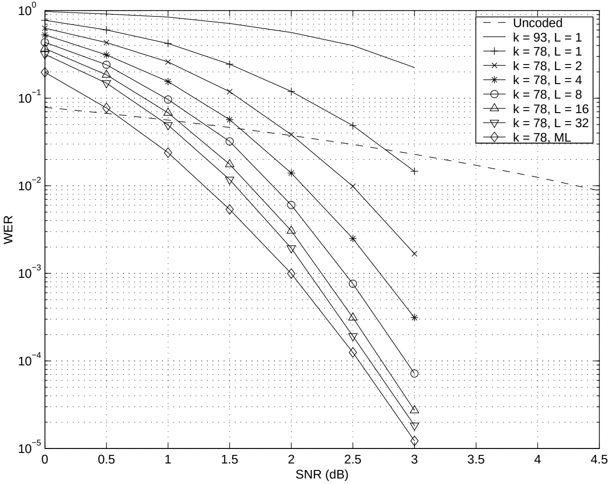


BSC_p with p=0.25 and g=0.5

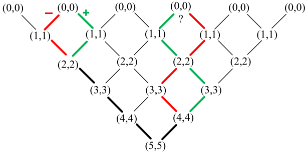
How can we improve recursive decoding?

1. Use ML-decoding on the end nodes of order $r \geq 1$ if possible;
2. Eliminate the weakest channels (bits) and consider subcodes;
3. Employ lists of code candidates in intermediate decoding steps.

Example. Optimized subcodes with end ML-decoding and intermediate lists in RM code $\mathcal{R}(3,8)$ [Dumer, Shabunov '2000]
$[n=256, k=93]$ RM code and its $[256,78]$ subcode



RM and Polar codes: $\mathcal{R}(m,m)$ has 2^m paths; one inform. bit per path.



Polarization Theorem [Arikan '2009]. Recursive decoding of $\mathcal{R}(m,m)$ on a memoryless channel W with (symmetric) capacity $I(W)$ gives

$$\lim_{n\to\infty}[\text{the fraction of good paths with } P_e \to 0] \quad \to \quad I(W)$$
$$\lim_{n\to\infty}[\text{the fraction of bad paths with } P_e \nrightarrow 0] \quad \to \quad 1 - I(W).$$

The fraction of good paths reaches the symmetric capacity $I(W)$
(i.e. the Shannon's capacity for equiprobable alphabets).

Shortcomings of polar codes:

1. Many good paths have slowly declining prob. $P_e \sim \exp\{-c\sqrt{n}\}$.
2. Good paths lack explicit description on a given error channel.

Consider both probability offsets $g_i = 1 - 2p_i$ and likelihoods $h_i = p_i/(1 - p_i)$.

To estimate performance of a path,we take $\mathbf{c} = \mathbf{1}^n$ and assume that previous paths give correct $v_i = 1$. Then recalculations for $h_{i,(i)}$ give

$$h_i^{\mathbf{v}} = \frac{h_i + h_{(i)}}{1 + h_i h_{(i)}}, \qquad h^{\mathbf{u}} = h_i h_{(i)}$$

Any path $\xi = (\xi_1, \ldots, \xi_m)$ derives $h_i^{\mathbf{v}}$ if $\xi_i = 1$ and $h_i^{\mathbf{u}}$ if $\xi_i = 0$.

We bound error rate via the expectation $\mathbb{E}h^{\lambda}(\xi)$ of $h^{\lambda}(\xi)$. Then

$$\Pr\{h(\xi) > 1\} \leq \min_{\lambda > 0} \mathbb{E}h^{\lambda}(\xi)$$

Theorem 1. For any subpaths $\bar{\xi} = (\xi_1, \ldots, \xi_i)$, $\bar{\xi}_{\mathbf{v}} = \left(\bar{\xi}, 1\right)$ and $\bar{\xi}_{\mathbf{u}} = \left(\bar{\xi}, 0\right)$

$$\begin{cases} \mathbb{E}\left(h_{\mathbf{u}}^{\lambda}\right) < \mathbb{E}\left(h^{\lambda}\right) \leq \mathbb{E}\left(h_{\mathbf{v}}^{\lambda}\right) \\ \mathbb{E}\left(h_{\mathbf{u}}^{\lambda}\right) + \mathbb{E}\left(h_{\mathbf{v}}^{\lambda}\right) \leq 2\mathbb{E}\left(h^{\lambda}\right) \end{cases} \qquad \text{if } \lambda \in (0, 1]$$

Theorem 2. For two neighbor-paths $\xi_{\mathbf{uv}} = \left(\bar{\xi}, 0, 1\right)$ and $\xi_{\mathbf{vu}} = \left(\bar{\xi}, 1, 0\right)$

$$\mathbb{E}h_{\mathbf{uv}}^{\lambda} \leq \mathbb{E}h_{\mathbf{vu}}^{\lambda} \quad \lambda \in [0, 1]$$

Examples. Consider code $\mathcal{R}(r,m)$ with $r \sim m/2$ and rate $R \in (0,1)$. We use it on a BSC_p with offset $g = 1-2p$. To find the worst path ξ^*, we can replace any $\mathbf{uv}$-segment with $\mathbf{vu}$ on any path ξ.

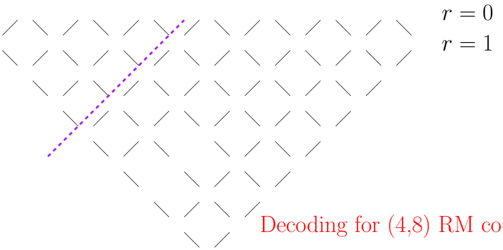
Full Paths: let decoding end on single bits $\mathcal{R}(0,0)$. Then $\xi^* = 1^r 0^{m-r}$.

Prefix 1^r gives offset $(1-2p)^{2^r} \sim \exp\{-2^{r+1}p\}$. Suffix 0^{m-r} is a repetition code $[d = 2^{m-r}, 1, d]$. Its error rate $P(\xi) \to 0$ if

$$\exp\{-2^{r+1}p\} \lesssim d/2, \; pn \lesssim (d \ln d)/4.$$

1-Truncated paths: let decoding end on biorthogonal end nodes $\mathcal{R}(1,\ell)$.

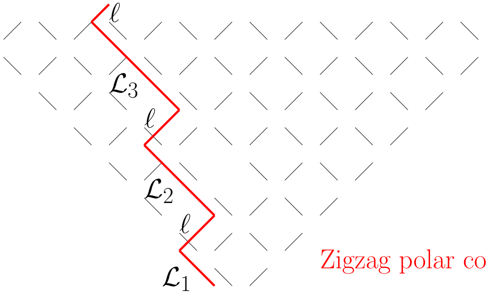
Then $\xi^* = 1^{r-1}0^{m-r}$ and we obtain similar estimate $pn \lesssim (d \ln d)/2$.



Decoding for (4,8) RM code

Codes of length 2^m with rate $R \to 1$ and error probability $p \sim 1/m$.

Bound all paths by a zigzag with $\mathbf{v}$-paths of increasing length $\mathcal{L}_i = 2^{i-2} \log m$ and $\mathbf{u}$–paths with $\ell = 3$. $\mathcal{L}_1$-path has length $(\log m)/2$. It gives offset $(1 - 2/m)^{\mathcal{L}_1} \sim 1 - 2/\sqrt{m}$. Then ℓ-path is a repetition code of length $d = 8$. It has error rate of order $(1/\sqrt{m})^{d/2} = m^{-2}$. The next section of $\mathcal{L}_2$ and ℓ gives error rate of order m^{-4} and so on. Thus, $P(\xi) \to 0$. The overall rate is $R \geq 1 - cH(6/\log m) \to 1$. Generally, codes have $R = 1 - c_1 H(c_2 p)$

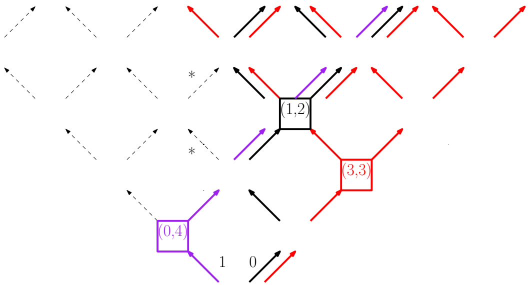


Zigzag polar code from (8,8) RM code

Codes $\mathcal{R}(r,m)$ with $r \succ m/2$ have rate $R \to 1$, and include polar codes. RM codes optimize path' choice w.r.t. distance d, and polar codes do it w.r.t. error probability P_e. This choice depends on specific decoding.

Consider a subcode C_ξ of $\mathcal{R}(m,m)$ that takes one path $\xi = (i_1, ..., i_{m-\ell})$ of length $m-\ell$ and extends it with some end node $R_\xi(s,\ell)$ of dimension k_ξ. C_ξ has distance $d_\xi = 2^{m-s-wt(\xi)}$. To enhance polar codes, we use some subset of paths T and ML-decode the corresponding end nodes $R_\xi(s,\ell)$. Then we obtain code $C(m,T) = \cup_{\xi\in T} C(\xi)$ with various end nodes.

Lemma. $C(m,T)$ has $n=2^m$, $k(m,T) = \sum_{\xi\in T} k_\xi$, $d(m,T) = \min_{\xi\in T} d_\xi$.



Paths: (0,0)/RM(3,3), (0,1,0)/RM(1,2), and 1/RM(0,4)

Let codes $C_\ell(m,T)$ have all end nodes $R_\xi(s,b)$ with parameters $b \le \ell$. For small ℓ, ML-decoding of $R_\xi(s,b)$ only slightly increases complexity.

Lemma. Recursive decoding of any code $C_\ell(m,T)$ with ML-decoding of end nodes has complexity $\prec n^{1+1/\ln m}$ if $\ell/\log_2 m < 1$ (codes C_0 give $n \ln n$).

ML-decoding of short nodes $R_\xi(s,b)$ also retains the asymptotic BER.
Lemma. For $m \to \infty$, polar codes $C_0(m,T)$ of rate R yield some seq-ce of polarized codes $C_\ell(m,T')$ of similar rate $\rho \to R$ if $\ell \le \log_2 m$.

Node $R_\xi(s,\ell)$ includes all paths (ξ,η) with a common prefix ξ and all suffixes η of length ℓ and weight $\ge s$. ML-decoding of $R_\xi(s,\ell)$ replaces recursive decoding of various subpaths η with the single subpath $\mathbf{0}^{\ell-s}$ that passes the upgraded u-channel $\ell - s$ times.

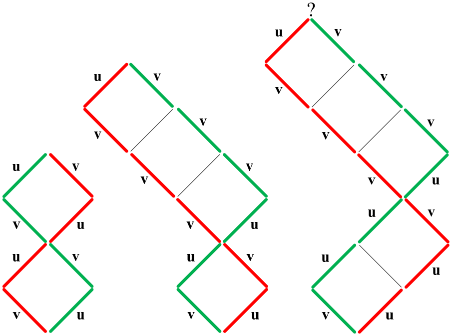
Lemma. Recursive decoding of code $C(\xi)$ with ML-decoding of its end node $R(s,\ell)$ has error probability $P(\xi) \le 2^{k_\xi} p\left(\overline{\xi}\right)$, where $p\left(\overline{\xi}\right)$ is the error probability of the extended path $\overline{\xi} = \xi, \mathbf{0}^{\ell-s}$.

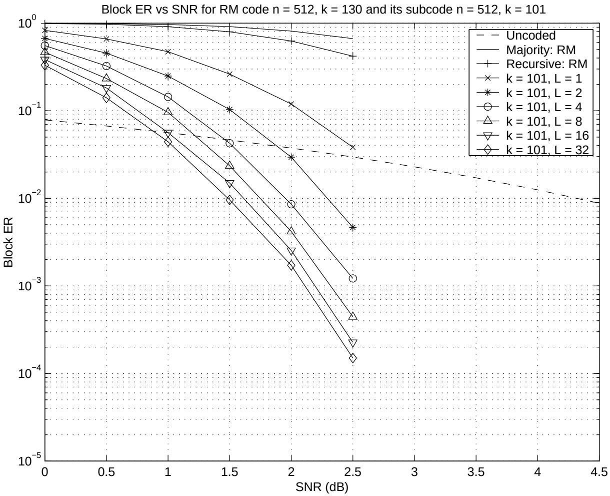
Open questions

For $\epsilon > 0$ and $n = 2^m$ let $I(p, m, \epsilon)$ be a set of $2^{n(C-\epsilon)}$ recursive paths that achieve $P_e \to 0$ on BSC$_p$ of capacity $C = 1 - H(p)$ as $m \to \infty$

We call the sets $I(p, m, \epsilon)$ weakly embedded if $I(q, m, \epsilon) \subset I(p, m, \epsilon)$ for all $(p, q) : p < q < 1/2$ as $m \to \infty$. They are strongly embedded if $I(q, m, \epsilon)$ is the subset of the "best" channels in $I(p, m, \epsilon)$.

1. Are subsets $I(p, m, \epsilon)$ weakly embedded for all $p < q < 1/2$?
2. Can subsets $I(p, m, \epsilon)$ be strongly embedded for some channels?





Block ER vs SNR for RM code n = 512, k = 130 and its subcode n = 512, k = 101
Block ER
SNR (dB)
Uncoded
Majority: RM
Recursive: RM
k = 101, L = 1
k = 101, L = 2
k = 101, L = 4
k = 101, L = 8
k = 101, L = 16
k = 101, L = 32

Decoding performance of RM codes of fixed rate R (order $r \sim m/2$)

Majority Decoding [Reed '54, Krichevskiy '70]

Corrects $\simeq (d \ln d)/4$ errors Complexity $O(nk)$

Distance-based recursive decoding
[Litsyn '88, Kabatyanski '90, Schnabl-Bossert '95]

Corrects $d/2$ errors
(up to $\simeq (d \ln d)/4$ errors) Complexity $O(n \log n)$

Probabilistic recursive decoding [Dumer '99]

Corrects $\simeq (d \ln d)/2$ errors
($\simeq n$(1-o(1))/2 for const order r) Complexity $O(n \log n)$

Probabilistic recursive decoding is analyzed as follows.

1. Separate decoding for different information bits;
2. End recursion on biorthogonal codes instead of repetition codes;
3. Find and eliminate the most error-prone information bits.