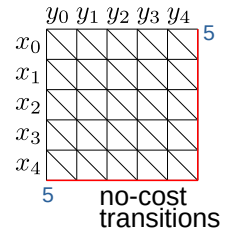


Definition of inner code

β : code length
 ν : number of codes
 $\mathcal{C}_i \subset \Sigma^\beta$: code ($i \in [\nu]$)
 $\mathbf{b} = (b_0, \dots, b_{\nu-1})$: bit-width $|\mathcal{C}_i| = 2^{b_i}$
 $R_1 = \frac{1}{2n\beta} \sum_{i=0}^{\nu-1} b_i$: inner code rate

Metric

$\mathbf{x} = (x_0, \dots, x_{\beta-1}) \in \mathbb{B}^\beta$
 $\mathbf{y} = (y_0, \dots, y_{\beta-1}) \in \mathbb{B}^\beta$
 metric $d(\mathbf{x}, \mathbf{y})$:
 insertion: (insert 0 or 1) + (delete the last bit)
 deletion: (delete one bit) + (append 0 or 1)
 substitution: invert one bit



Construction from binary code

$\mathbf{b}' = (b'_0, \dots, b'_{\nu-1}) \in [\beta + 1]^\nu$ ($b_i = b'_i + \beta$)

even β : $2^{b'_i} \leq \binom{\beta}{\beta/2}$

odd β : $2^{b'_i} \leq \binom{\beta}{(\beta-1)/2} = \binom{\beta}{(\beta+1)/2}$

[computer search]

input: β, b', t

output: $\mathcal{B} \subset \mathbb{B}^\beta$

constraint: $|\mathcal{B}| = 2^{b'}$, $\forall \mathbf{u} \in \mathcal{B}, w(\mathbf{u}) = t$

objective:

(1) maximize: $\min_{\mathbf{x}, \mathbf{y} \in \mathcal{B}} d(\mathbf{x}, \mathbf{y})$ ($\mathbf{x} \neq \mathbf{y}$)

(2) maximize:
(+ total sum?)

β	t	$\binom{\beta}{t}$	b'					
			1	2	3	4	5	6
2	{1}	2	$\frac{2}{2}$	—	—	—	—	—
3	{1, 2}	3	$\frac{2}{3}$	—	—	—	—	—
4	{2}	6	$\frac{2}{6}$	$\frac{4}{6}$	—	—	—	—
5	{2, 3}	10	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{8}{10}$	—	—	—
6	{3}	20	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{8}{20}$	$\frac{16}{20}$	—	—
7	{3, 4}	35	$\frac{2}{35}$	$\frac{4}{35}$	$\frac{8}{35}$	$\frac{16}{35}$	$\frac{32}{35}$	—
8	{4}	70	$\frac{2}{70}$	$\frac{4}{70}$	$\frac{8}{70}$	$\frac{16}{70}$	$\frac{32}{70}$	$\frac{64}{70}$
9	{4, 5}	126	$\frac{2}{126}$	$\frac{4}{126}$	$\frac{8}{126}$	$\frac{16}{126}$	$\frac{32}{126}$	$\frac{64}{126}$

Evaluation of inner code

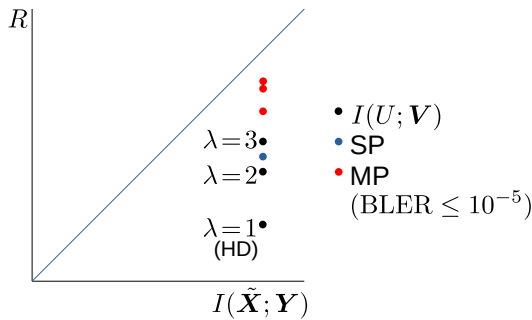
channel input: $\mathbf{u} \in \mathbb{F}_{2^{b(i, \nu)}}$

channel output: $\mathbf{p} \in \mathbb{P}_{2^{b(i, \nu)}}$ (APP from inner code)

(a) HD+H(p): $\mathbf{v} = (v, Q_\lambda(H(\mathbf{p})))$
 Q_λ : λ -level quantization

(b) list output:
 (list size λ) $\mathbf{v} = (v_0, \dots, v_{\lambda-1})$
 $= \arg \max(\mathbf{p}, \lambda) \in \mathbb{F}_{2^{b(i, \nu)}}^\lambda$

mutual info.: $I(U; \mathbf{V})$



Concatenated coding

channel alphabet: $\Sigma = \{A, T, G, C\}$

outer code: $\mathcal{C}_o \subset \mathbb{F}_{2^b}^{n_o}$

n_o : code length

b : symbol size (bits)

inner code: $\mathcal{C}_i \subset \Sigma^\beta$ ($i \in [\nu]$, $b < 2\beta$)

β : code length (symbols)

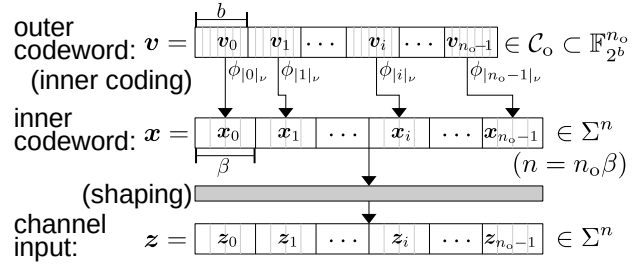
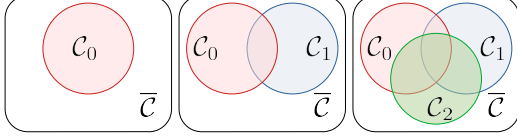
ν : number of code books

$|\mathcal{C}_i| = 2^b$: number of codewords

$R = b/2\beta$: inner code rate

$\phi_i: \mathbb{F}_{2^b} \rightarrow \mathcal{C}_i$: encoding function

Σ^β ($\nu = 1$) ($\nu = 2$) ($\nu = 3$)



params:

β	b
3	4, 5
4	6, 7
5	8, 9
6	10, 11
$\nu = 2$	

$$\text{prior: } p(x_i) = \begin{cases} 1/2^b & (x_i \in \mathcal{C}_{i|\nu}) \\ 0 & (x_i \notin \mathcal{C}_{i|\nu}) \end{cases}$$

Constraint

RL-LB constrained coding ($\ell; w, \delta$):

* run length coding: $f_R(z) \leq \ell$

* local GC-balance: $|f_B(z_{i+w-1}^{i+w-1})| \leq \delta$
($\forall i \in [n-w]$)

* synchronization: MIR?

vector over Σ : $u = (u_0, \dots, u_{n-1}) \in \Sigma^n$

weight: $w(u)_x = |\{i \in [n] \mid u_i = x\}|$ ($x \in \Sigma$)

balance: $f_B(u) = w_G(u) + w_C(u) - w_A(u) - w_T(u) \in [-n, n]$

max run length: $f_R(u) \in [1, n]$

Inner codebook

$$\mathcal{C}_i = \{(\phi(x_0 y_0), \dots, \phi(x_{\beta-1} y_{\beta-1})) \mid x \in \mathcal{B}_i, y \in \mathbb{B}^\beta\}$$

$$x = (x_0, \dots, x_{\beta-1})$$

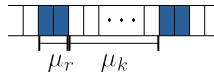
$$y = (y_0, \dots, y_{\beta-1}) \quad (i \in [\nu])$$

$$\mathcal{B}_i \subset \mathbb{B}^\beta$$

$$|\mathcal{C}_i| = 2^\beta |\mathcal{B}_i| = 2^b, \quad |\mathcal{B}_i| = 2^{b-\beta}$$

v	$\phi(v)$
00	A
01	T
10	G
11	C

Marker



$$\text{rate: } R_i = \frac{\mu_k}{\mu_k + \mu_r}$$

Rate

$$R = R_o R_i R_c \text{ [bits/symbol]}$$

$$R_o \leq 1 - \text{cross entropy: inner APP}$$

$$R_i = \text{inner code rate}$$

$$R_c \leq \text{bound of } (\ell; w, \delta) \text{ constraint}$$

$$\text{bound for given } P(X)$$

(example)

	β	ν	R_i	ℓ	δ	b_i	\mathcal{B}_i
1	3.4.2	3	2	4/6	2	2	4 \mathcal{B}_0
							4 \mathcal{B}_1
2	3.5.1	3	1	5/6	3	*	5 \mathcal{B}_0
3	4.5.2	4	2	5/8	2	2	5 \mathcal{B}_0
							5 \mathcal{B}_1
4	4.6.1a	4	1	6/8	2	2	6 \mathcal{B}_0
5	4.6.1b	4	1	6/8	4	*	6 \mathcal{B}_0
6	4.7.1	4	1	7/8	4	*	7 \mathcal{B}_0
7	5.7.2	5	2	7/10	5	5	7 \mathcal{B}_0
							7 \mathcal{B}_1
9	5.8.2	5	2	8/10	3	3	8 \mathcal{B}_0
							8 \mathcal{B}_1
10	5.9.1	5	1	9/10	4	*	9 \mathcal{B}_0
8	5.a.2	5	2	15/20	3	3	8 \mathcal{B}_0
							7 \mathcal{B}_1

(ex2b)

β	3	3	4	4	4	4	5	5	6
b	4	5	5	6	6	7	8	9	10
ν	2	1	2	1	1	1	2	1	1
R_i	0.67	0.83	0.63	0.75	0.75	0.88	0.80	0.90	0.83
				a	b				
$(\ell; w, \delta)$	(2; 10, 2)	(3; *, *)	(2; 10, 2)	(2; 10, 2)	(4; *, *)	(4; *, *)	(3; 10, 2)	(4; *, *)	

* (multi-read)

NB-IDS channel

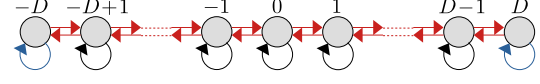
block length: n
 alphabet: Σ ($|\Sigma| = q$)
 error prob.: p_{id} ($< \frac{1}{2}$) (ins/del)
 p_s (sub)
 input: $\mathbf{x} = (x_0, \dots, x_{n-1}) = \Sigma^n$
 output: $\mathbf{y} = (y_0, \dots, y_{n'-1}) = \Sigma^{n'}$
 $(n-D \leq n' \leq n+D)$

[transmission]

1) drift vector: $\mathbf{d} = (d_0, \dots, d_{n-1}, d_n) \in \mathcal{D}^{n+1}$
 2) intermediate vector: $\mathbf{z} = (z_0, \dots, z_{n'-1}) \in \Sigma^{n'}$
 $z_j = x_i$ ($j \in [i + d_i, i + d_{i+1}]$)
 $i \in [n], n' = n + d_n$
 3) output vector: $p(y_i|z_i) = \begin{cases} 1 - p_s & (y_i = z_i) \\ \frac{p_s}{q-1} & (y_i \neq z_i) \end{cases}$

max drift: D
 set of drift values: $\mathcal{D} = \{-D, \dots, 0, \dots, D\}$
 drift vector: $\mathbf{d} = (d_0=0, \dots, d_{n-1}, d_n) \in \mathcal{D}^{n+1}$

$$p(d_{i+1}|d_i) = \begin{cases} 1 - 2p_{id} & (d_{i+1} = d_i, |d_i| < D) \rightarrow \\ 1 - p_{id} & (d_{i+1} = d_i, |d_i| = D) \rightarrow \\ p_{id} & (|d_{i+1} - d_i| = 1, |d_{i+1}| \leq D) \rightarrow \\ 0 & (\text{otherwise}) \end{cases}$$



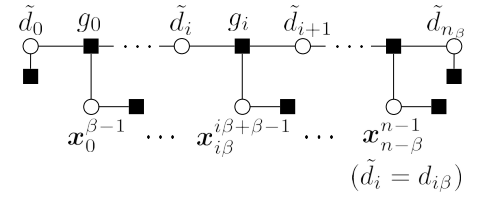
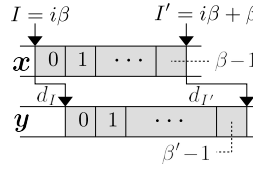
APP by FBA

$$p(\mathbf{x}, \mathbf{y}, \mathbf{d}) = p(\mathbf{y}|\mathbf{x}, \mathbf{d})p(\mathbf{x}, \mathbf{d}) = p(\mathbf{y}|\mathbf{x}, \mathbf{d})p(\mathbf{x})p(\mathbf{d})$$

$$= p(d_0) \prod_{i=0}^{n-1} p\left(\mathbf{y}_{i+d_i}^{i+d_{i+1}} \middle| x_i, d_i, d_{i+1}\right) p(d_{i+1}|d_i) p(x_i)$$

segment level synchronization: $n = n_\beta \beta$

$$= p(d_0) \prod_{i=0}^{n_\beta-1} \frac{p\left(\mathbf{y}_{I+d_I}^{I'-1+d_{I'}} \middle| \mathbf{x}_I^{I'-1}, d_I, d_{I'}\right) p(d_{I'}|d_I) p\left(\mathbf{x}_I^{I'-1}\right)}{(I = i\beta, I' = i\beta + \beta)}$$

[LUT] (for a large D)

$$G\left(\beta', \mathbf{y}_0^{\beta'-1}\right) \text{ list of } \left[\mathbf{x}_0^{\beta-1}, p_0\left(\mathbf{y}_0^{\beta'-1} \middle| \mathbf{x}_0^{\beta-1}, 0, d'\right) p_0(d'|0) (> p_{th}) \right]$$

$$p_\alpha\left(\mathbf{y}_{\alpha'}^{\beta'-1} \middle| \mathbf{x}_\alpha^{\beta-1}, d, d'\right) p_\alpha(d'|d) = \begin{matrix} (d' = \beta' - \beta) \\ (d = \alpha' - \alpha) \end{matrix}$$

$$\sum_{\delta \in \{-1, 0, 1\}} p\left(\mathbf{y}_{\alpha'+\delta}^{\alpha'+\delta} \middle| \mathbf{x}_\alpha, d, d + \delta\right) p(d + \delta|d)$$

$$\times p_{\alpha+1}\left(\mathbf{y}_{\alpha'+\delta+1}^{\beta'} \middle| \mathbf{x}_{\alpha+1}^{\beta-1}, d + \delta, d'\right) p_{\alpha+1}(d'|d + \delta)$$

Gtable:
 int beta
 double Pid
 double Ps
 double Pth
 LIST[Len][Y]

LIST:
 int num
 XP[num]

XP:
 int x
 float p

[run-length of random sequence]

prob. of run-length ℓ : $p(\ell) = \left(\frac{1}{q}\right)^{\ell-1} \left(\frac{q-1}{q}\right) = \frac{q-1}{q^\ell}$

expectation: $E(\ell) = \sum_{\ell=1}^n \ell p(\ell) = \sum_{\ell=1}^n \frac{(q-1)\ell}{q^\ell}$
 $= (q-1) \sum_{\ell=1}^n \frac{\ell}{q^\ell}$
 $= \frac{q}{q-1} \left(1 - \frac{1}{q^n}\right) - \frac{n}{q^n}$

[local GC-balance of random sequence]

window size: w

prob. binary vector having weight t : $p_w(t) = \binom{w}{t} / 2^w$

absolute GC-balance: $\sum_{t=0}^w p_w(t) |w - 2t|$

[rate upper bound]

run length: constraint graph(?)

local GCB: $\frac{1}{w} \log_2 \left(2^w \sum_{t \in \mathcal{T}} \binom{w}{t} \right) = 1 + \frac{1}{w} \log_2 \left(\sum_{t \in \mathcal{T}} \binom{w}{t} \right)$
 $(\mathcal{T} = \{t \mid |w - 2t| \leq \delta\})$

RL+LGCB: $\frac{1}{w} \log_2 \left| \left\{ \mathbf{u} \in \Sigma^w \mid |f_B(\mathbf{u})| \leq \delta, f_R(\mathbf{u}) \leq \ell \right\} \right|$
 $(\ell; w, \delta)$

Concatenated coding

channel alphabet: $\Sigma = \{A, T, G, C\}$

outer code: $\mathcal{C}_o \subset \mathbb{F}_{2^b}^{n_o}$

n_o : code length

b : symbol size (bits)

inner code: $\mathcal{C}_i \subset \Sigma^\beta$ ($i \in [\nu]$, $b < 2\beta$)

β : code length (symbols)

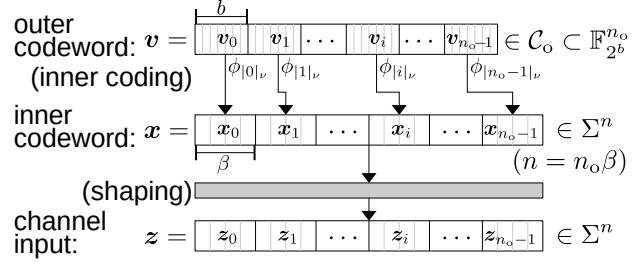
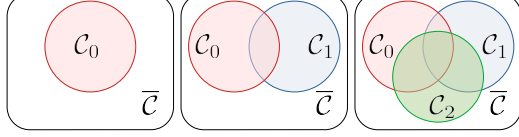
ν : number of code books

$|\mathcal{C}_i| = 2^\beta$: number of codewords

$R = b/2\beta$: inner code rate

$\phi_i: \mathbb{F}_{2^b} \rightarrow \mathcal{C}_i$: encoding function

Σ^β ($\nu = 1$) ($\nu = 2$) ($\nu = 3$)



params:

β	b
3	4, 5
4	6, 7
5	8, 9
6	10, 11

prior: $p(x_i) = \begin{cases} 1/2^\beta & (x_i \in \mathcal{C}_{i|\nu}) \\ 0 & (x_i \notin \mathcal{C}_{i|\nu}) \end{cases}$

vector over Σ : $\mathbf{u} = (u_0, \dots, u_{n-1}) \in \Sigma^n$

weight: $w(\mathbf{u})_x = |\{i \in [n] \mid u_i = x\}|$ ($x \in \Sigma$)

balance: $f_B(\mathbf{u}) = w_G(\mathbf{u}) + w_C(\mathbf{u}) - w_A(\mathbf{u}) - w_T(\mathbf{u}) \in [-n, n]$

max run length: $f_R(\mathbf{u}) \in [1, n]$

Constraint

RL-LB constrained coding ($\ell; w, \delta$):

* run length coding: $f_R(\mathbf{z}) \leq \ell$

* local GC-balance: $|f_B(\mathbf{z}_{i+w-1}^{i+w-1})| \leq \delta$
($\forall i \in [n-w]$)

* synchronization: MIR?

[example 1]

forbidden set $\bar{\mathcal{C}} \subseteq \Sigma^\beta \setminus \bigcup_{i \in [\nu]} \mathcal{C}_i$ ($|\bar{\mathcal{C}}| \leq 2^{2\beta} - 2^\beta$)

$\bar{\mathcal{C}}$ (empty)

$\bar{\mathcal{C}}_{*,\lambda} = \{\mathbf{u} \in \Sigma^\beta \mid f_R(\mathbf{u}) \geq \lambda\}$ (RL)

$\bar{\mathcal{C}}_{\omega,*} = \{\mathbf{u} \in \Sigma^\beta \mid |f_B(\mathbf{u})| \geq \omega\}$ (LB)

$\bar{\mathcal{C}}_{\omega,\lambda} = \bar{\mathcal{C}}_{\omega,*} \cup \bar{\mathcal{C}}_{*,\lambda}$ (both)

[example 2a] $\beta = 4, b = 7$

$\mathcal{C}_i = \{(\phi(x_0 y_0), \dots, \phi(x_3 y_3)) \mid (x_0, \dots, x_3) \in \mathcal{B}_i, (y_0, \dots, y_3) \in \mathbb{B}^4\}$

$\mathcal{B}_0: 0011, 0101, 0110, 1001, 1010, 1100, \mathbf{0010, 0100}$

$\mathcal{B}_1: 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, 00111$

$|\mathcal{C}_i| = 8 \times 16 = 128 = 2^\beta$

\mathbf{v}	$\phi(\mathbf{v})$
00	A
01	T
10	G
11	C

[example 2b] $\beta = 5, b = 8$

$\mathcal{C}_i = \{(\phi(x_0 y_0), \dots, \phi(x_4 y_4)) \mid (x_0, \dots, x_4) \in \mathcal{B}_i, (y_0, \dots, y_4) \in \mathbb{B}^5\}$

$\mathcal{B}_0: 00011, 00101, 00110, 01001, 01010, 01100, 10001, 10010, 10100, 11000$

$\mathcal{B}_1: 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, 00111$

$|\mathcal{C}_i| = 8 \times 32 = 256 = 2^\beta$

$\mathbf{a} = (a_0, \dots, a_4), \mathbf{c} = (c_0, \dots, c_4) \in \mathcal{B}_0$

$\mathbf{b} = (b_0, \dots, b_4), \mathbf{d} = (d_0, \dots, d_4) \in \mathcal{B}_1$

RL: max RL in \mathbf{a} and \mathbf{b} : 3 (by definition)

max RL in (\mathbf{a}, \mathbf{b}) : 3 (proof 1)

max RL in (\mathbf{b}, \mathbf{c}) : 3 (proof 2)

(proof 1) length of the last run of \mathbf{a} : 1 (run of 1)

2 (run of 0s)

length of the first run of \mathbf{b} : 2 (run of 1s)

1 (run of 0)

(proof 2) length of the last run of \mathbf{b} : 1 (run of 0)

2 (run of 1s)

length of the first run of \mathbf{c} : 2 (run of 0s)

1 (run of 1)

LB: even $w(\geq 10)$

$(\mathbf{a}_t^4, \mathbf{b}, \mathbf{c}_0^{t-1})$:

$$\frac{1}{w} \log_2 \left(\sum_{t \in \mathcal{T}} \left(2^w \binom{w}{t} - \rho_{\ell,w,t} \right) \right)$$

$$\rho_{\ell,w,t} = |\{\boldsymbol{u} \in \Sigma^w | w_{\mathsf{G}}(\boldsymbol{u}) + w_{\mathsf{C}}(\boldsymbol{u}) = t, f_{\mathsf{R}}(\boldsymbol{u}) > \ell\}|$$

$(\beta = 3, b = 4)$
 $\mathcal{B}_0 : 010, 100$
 $\mathcal{B}_1 : 101, 110$

$(\beta = 3, b = 5)$
 $\mathcal{B}_0 : 001, 010, 100, \textcolor{red}{101}$
 $\mathcal{B}_1 : 110, 101, 011, \textcolor{red}{010}$

$(\beta = 4, b = 6)$
 $\mathcal{B}_0 : 0101, 0110, 1001, 1010$
 $\mathcal{B}_1 : 0101, 0110, 1001, 1010$

$(\beta = 4, b = 7)$ [ex2a]
 $\mathcal{B}_0 : 0011, 0101, 0110, 1001, 1010, 1100, \textcolor{red}{0010, 0100}$
 $\mathcal{B}_1 : 0011, 0101, 0110, 1001, 1010, 1100, \textcolor{red}{1101, 1011}$

$(\beta = 5, b = 8)$ [ex2b]
 $\mathcal{B}_0 : 00101, 00110, 01001, 01010, 01100, 10001, 10010, 10100$
 $\mathcal{B}_1 : 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011$

$(\beta = 5, b = 9)$
 $\mathcal{B}_0 : 00101, 00110, 01001, 01010, 01100, 10001, 10010, 10100, \dots$
 $\mathcal{B}_1 : 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, \dots$

position	max RL	pattern
$(a_0, a_1, a_2, a_3, a_4)$	3	10001
$(a_1, a_2, a_3, a_4; b_0)$	3	*100; 0
$(a_2, a_3, a_4; b_0, b_1)$		
$(a_3, a_4; b_0, b_1, b_2)$		
$(a_4; b_0, b_1, b_2, b_3)$		
$(b_0, b_1, b_2, b_3, b_4)$	3	01110
$(b_1, b_2, b_3, b_4; c_0)$		
$(b_2, b_3, b_4; c_0, c_1)$		
$(b_3, b_4; c_0, c_1, c_2)$		
$(b_4; c_0, c_1, c_2, c_3)$		