Concatenated coding

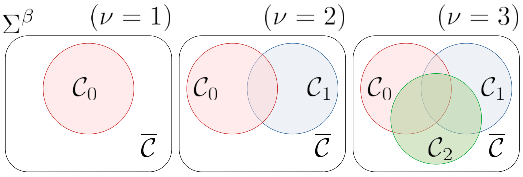
channel alphabet: $\Sigma = \{\mathsf{A},\mathsf{T},\mathsf{G},\mathsf{C}\}$

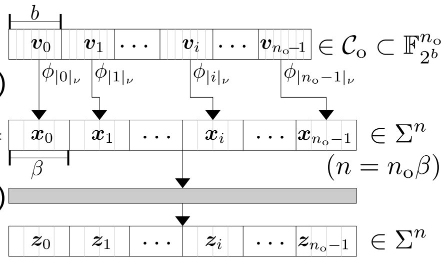
outer code: $\mathcal{C}_\mathrm{o} \subset \mathbb{F}_{2^b}^{n_\mathrm{o}}$

- n_o : code length
- b : symbol size (bits)

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

- β : code length (symbols)
- ν : number of code books
- $|\mathcal{C}_i| = 2^b$: number of codewords
- $R = b/2\beta$: inner code rate
- $\phi_i : \mathbb{F}_{2^b} \to \mathcal{C}_i$: encoding function





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

params:

β	b
3	4, 5
4	6, 7
5	8, 9
6	10, 11

$\nu = 2$

Constraint

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

* run length coding: $f_\mathrm{R}(\boldsymbol{z}) \le \ell$
* local GC-balance: $\left|f_\mathrm{B}\left(\boldsymbol{z}_i^{i+w-1}\right)\right| \le \delta \quad (\forall i \in [n-w\rangle)$
* synchronization: MIR?

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

weight: $w(\boldsymbol{u})_x = |\{i \in [n\rangle \,|\, u_i = x\}| \quad (x \in \Sigma)$

balance: $f_\mathrm{B}(\boldsymbol{u}) = w_\mathsf{G}(\boldsymbol{u}) + w_\mathsf{C}(\boldsymbol{u}) - w_\mathsf{A}(\boldsymbol{u}) - w_\mathsf{T}(\boldsymbol{u}) \in [-n, n]$

max run length: $f_\mathrm{R}(\boldsymbol{u}) \in [1, n]$

Inner codebook

$\mathcal{C}_i = \left\{(\phi(x_0 y_0), \ldots, \phi(x_{\beta-1} y_{\beta-1})) \middle| \boldsymbol{x} \in \mathcal{B}_i, \boldsymbol{y} \in \mathbb{B}^\beta\right\}$

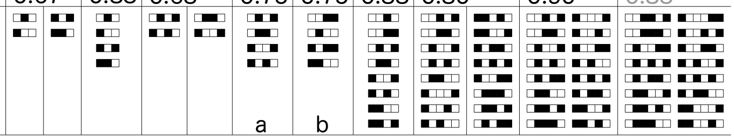
$\boldsymbol{x} = (x_0, \ldots, x_{\beta-1})$
$\boldsymbol{y} = (y_0, \ldots, y_{\beta-1}) \quad (i \in [\nu\rangle)$
$\mathcal{B}_i \subset \mathbb{B}^\beta$
$|\mathcal{C}_i| = 2^\beta |\mathcal{B}_i| = 2^b, \; |\mathcal{B}_i| = 2^{b-\beta}$

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

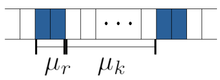
(ex2b)

(example)

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				



Marker



rate: $R_\mathrm{i} = \frac{\mu_k}{\mu_k + \mu_r}$

Rate

$R = R_\mathrm{o} R_\mathrm{i} R_\mathrm{c}$ [bits/symbol]

- $R_\mathrm{o} \le$ 1-cross entropy: inner APP
- $R_\mathrm{i} =$ inner code rate
- $R_\mathrm{c} \le$ bound of $(\ell; w, \delta)$ constraint
- $R_\mathrm{o} R_\mathrm{i} \le I(X;Y)$: bound for given $P(X)$

* (multi-read)

NB-IDS channel

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

[transmission]

1) drift vector: $\boldsymbol{d} = (d_0, \ldots, d_{n-1}, d_n) \in \mathcal{D}^{n+1}$
2) intermediate vector: $\boldsymbol{z} = (z_0, \ldots, z_{n'-1}) \in \Sigma^{n'}$

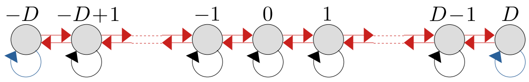
$$z_j = x_i \ (j \in [i + d_i, i + d_{i+1}]) \quad i \in [n\rangle, \ n' = n + d_n$$

3) output vector:

$$p(y_i|z_i) = \begin{cases} 1 - p_{\mathrm{s}} & (y_i = z_i) \\ \frac{p_{\mathrm{s}}}{q-1} & (y_i \neq z_i) \end{cases}$$

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

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



APP by FBA

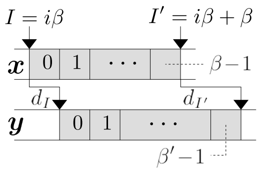
$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{d}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{d}) p(\boldsymbol{x}, \boldsymbol{d}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{d}) p(\boldsymbol{x}) p(\boldsymbol{d})$$

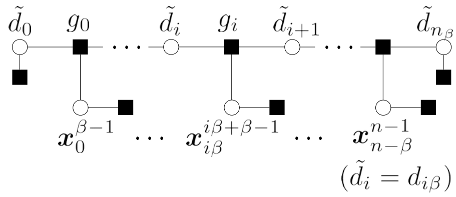
$$= p(d_0) \prod_{i=0}^{n-1} p\left(\boldsymbol{y}_{i+d_i}^{i+d_{i+1}} \middle| x_i, d_i, d_{i+1}\right) p(d_{i+1}|d_i) p(x_i)$$

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

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

$$(I = i\beta, I' = i\beta + \beta)$$





[LUT] (for a large D)

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

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

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

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

```
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: $\frac{1}{w}\log_2\left(2^w \sum_{t\in\mathcal{T}} \binom{w}{t}\right) = 1 + \frac{1}{w}\log_2\left(\sum_{t\in\mathcal{T}} \binom{w}{t}\right)$

$(\mathcal{T} = \{t \mid |w-2t| \le \delta\})$

RL+LGCB: $(\ell; w, \delta)$ $\frac{1}{w}\log_2\left|\left\{\boldsymbol{u}\in\Sigma^w \middle| |f_{\mathrm{B}}(\boldsymbol{u})| \le \delta, f_{\mathrm{R}}(\boldsymbol{u}) \le \ell\right\}\right|$

Concatenated coding

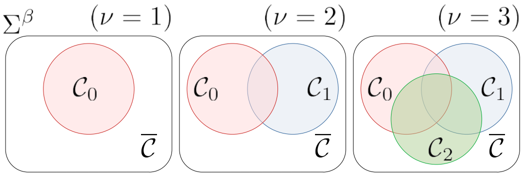
channel alphabet: $\Sigma = \{\mathsf{A}, \mathsf{T}, \mathsf{G}, \mathsf{C}\}$

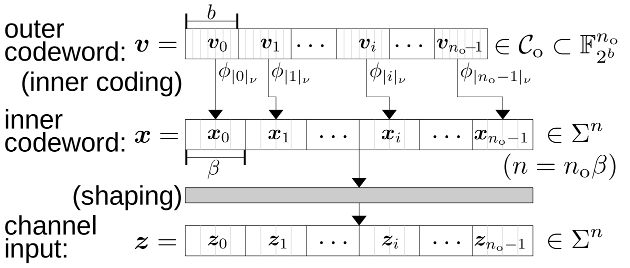
outer code: $\mathcal{C}_\mathrm{o} \subset \mathbb{F}_{2^b}^{n_\mathrm{o}}$

- n_o : code length
- b : symbol size (bits)

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

- β : code length (symbols)
- ν : number of code books
- $|\mathcal{C}_i| = 2^b$: number of codewords
- $R = b/2\beta$: inner code rate
- $\phi_i : \mathbb{F}_{2^b} \to \mathcal{C}_i$: encoding function





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

params:

β	b
3	$4, 5$
4	$6, 7$
5	$8, 9$
6	$10, 11$

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

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

balance: $f_\mathrm{B}(\boldsymbol{u}) = w_\mathsf{G}(\boldsymbol{u}) + w_\mathsf{C}(\boldsymbol{u}) - w_\mathsf{A}(\boldsymbol{u}) - w_\mathsf{T}(\boldsymbol{u}) \in [-n, n]$

max run length: $f_\mathrm{R}(\boldsymbol{u}) \in [1, n]$

Constraint

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

* run length coding: $f_\mathrm{R}(\boldsymbol{z}) \le \ell$
* local GC-balance: $\left|f_\mathrm{B}\left(\boldsymbol{z}_i^{i+w-1}\right)\right| \le \delta$ $\quad (\forall i \in [n - w\rangle)$
* synchronization: MIR?

[example 1]

forbidden set $\overline{\mathcal{C}} \subseteq \Sigma^\beta \setminus \cup_{i \in [\nu\rangle} \mathcal{C}_i \quad (|\overline{\mathcal{C}}| \le 2^{2\beta} - 2^b)$

- ϕ (empty)
- $\overline{\mathcal{C}}_{*,\lambda} = \left\{\boldsymbol{u} \in \Sigma^\beta \mid f_\mathrm{R}(\boldsymbol{u}) \ge \lambda\right\}$ (RL)
- $\overline{\mathcal{C}}_{\omega,*} = \left\{\boldsymbol{u} \in \Sigma^\beta \mid |f_\mathrm{B}(\boldsymbol{u})| \ge \omega\right\}$ (LB)
- $\overline{\mathcal{C}}_{\omega,\lambda} = \overline{\mathcal{C}}_{\omega,*} \cup \overline{\mathcal{C}}_{*,\lambda}$ (both)

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

$$\mathcal{C}_i = \left\{(\phi(x_0y_0), \ldots, \phi(x_3y_3)) \middle| (x_0, \ldots, x_3) \in \mathcal{B}_i, (y_0, \ldots, y_3) \in \mathbb{B}^4\right\}$$

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

$\mathcal{B}_1$:0011,0101,0110,1001,1010,1100,1101,1011

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

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

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

$$\mathcal{C}_i = \left\{(\phi(x_0y_0), \ldots, \phi(x_4y_4)) \middle| (x_0, \ldots, x_4) \in \mathcal{B}_i, (y_0, \ldots, y_4) \in \mathbb{B}^5\right\}$$

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

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

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

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

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

RL: max RL in $\boldsymbol{a}$ and $\boldsymbol{b}$: 3 (by definition)
max RL in $(\boldsymbol{a}, \boldsymbol{b})$: 3 (proof 1)
max RL in $(\boldsymbol{b}, \boldsymbol{c})$: 3 (proof 2)

(proof 1) length of the last run of $\boldsymbol{a}$: 1 (run of 1)
2 (run of 0s)
length of the first run of $\boldsymbol{b}$: 2 (run of 1s)
1 (run of 0)

(proof 2) length of the last run of $\boldsymbol{b}$: 1 (run of 0)
2 (run of 1s)
length of the first run of $\boldsymbol{c}$: 2 (run of 0s)
1 (run of 1)

LB: even $w (\ge 10)$

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

* (multi-read)

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

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

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

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

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

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

$(\beta=5,b=8)$ [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,...
$\mathcal{B}_1$: 11010,11001,10110,10101,10011,01110,01101,01011,...

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
$(b_1,b_2,b_3,b_4;c_0)$		
$(b_2,b_3,b_4;c_0,c_1)$		
$(b_3,b_4;c_0,c_1,c_2)$		
$(b_4;c_0,c_1,c_2,c_3)$		