Inner code

constrained coding
* run length
* GC-balance
* prior: p(x) for IDS

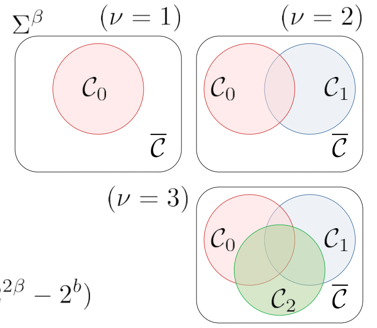
channel coding
* 4-ary IDS/asymmetric
* (multi-read)

inner code performance:
* run length distribution
* GC-weight distribution (sliding window)
* synchronization: cross entropy of APP(?)

alphabet: $\Sigma = \{\mathsf{A}, \mathsf{T}, \mathsf{G}, \mathsf{C}\}$
segment size: β [symbols]
segment: $\boldsymbol{u} = (u_0, \dots, u_{\beta-1}) \in \Sigma^\beta$

outer code: $\tilde{\mathcal{C}} \subset \mathbb{F}_{2^b}^n \quad (b < 2\beta)$

inner code: $\mathcal{C}_i \subset \Sigma^\beta \quad (i \in [\nu\rangle)$
(num CW) $|\mathcal{C}_i| = 2^b$
(rate) $R = b/2\beta$
(encoding) $\phi_i : \mathbb{F}_{2^b} \to \mathcal{C}_i$
(forbidden set) $\overline{\mathcal{C}} \subseteq \Sigma^\beta \setminus \cup_{i\in[\nu\rangle}\mathcal{C}_i \quad (|\overline{\mathcal{C}}| \leq 2^{2\beta} - 2^b)$



vector over Σ : $\boldsymbol{u} = (u_0, \dots, u_{n-1}) \in \Sigma^n$
weight: $w(\boldsymbol{u})_x = |\{i \in [n\rangle \mid u_i = x\}| \quad (x \in \Sigma)$
balance: $f_\mathrm{B}(\boldsymbol{u}) = w_\mathsf{G}(\boldsymbol{u}) + w_\mathsf{C}(\boldsymbol{u}) - w_\mathsf{A}(\boldsymbol{u}) - w_\mathsf{T}(\boldsymbol{u})$
max run length: $f_\mathrm{R}(\boldsymbol{u})$

[example 1]
forbidden set $\overline{\mathcal{C}}$:
ϕ (empty)
$\overline{\mathcal{C}}_{\omega,*} = \{\boldsymbol{u} \in \Sigma^\beta \mid |f_\mathrm{B}(\boldsymbol{u})| \geq \omega\}$ (GC-balance)
$\overline{\mathcal{C}}_{*,\lambda} = \{\boldsymbol{u} \in \Sigma^\beta \mid f_\mathrm{R}(\boldsymbol{u}) \geq \lambda\}$ (run length)
$\overline{\mathcal{C}}_{\omega,\lambda} = \overline{\mathcal{C}}_{\omega,*} \cup \overline{\mathcal{C}}_{*,\lambda}$ (both)

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

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

$\mathcal{B}_0$: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

[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:

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

[GC-balance of random sequence]

window size: ℓ

prob. binary vector having weight w:

$$p_\ell(w) = \binom{\ell}{w}/2^\ell$$

absolute GC-balance: $\sum_{w=0}^{\ell} p_\ell(w)\,|\ell - 2w|$

NB-IDS channel

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

[transmission]

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

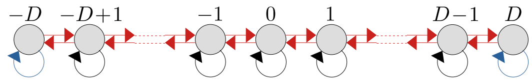
$$z_j = x_i \ (j \in [i + d_i, i + d_{i+1}]) \quad i \in [n\rangle,\ n' = n + d_n$$

3) output vector:

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

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

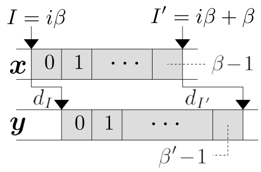
$$p(d_{i+1}|d_i) = \begin{cases} 1 - 2p_{\rm id} & (d_{i+1} = d_i, |d_i| < D) \\ 1 - p_{\rm id} & (d_{i+1} = d_i, |d_i| = D) \\ p_{\rm id} & (|d_{i+1} - d_i| = 1, |d_{i+1}| \le D) \\ 0 & (\text{otherwise}) \end{cases}$$

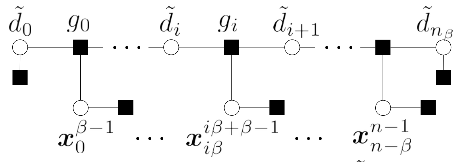


APP by FBA

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

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





$(\tilde{d}_i = d_{i\beta})$

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

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

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

[LUT] (for a large D)

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

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

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

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

```
Gtable:
  int beta
  double Pid
  double Ps
  double Pth
  LIST[len][y]
```

```
LIST:
  int num
  XP[num]
```

```
XP:
  int x
  float p
```

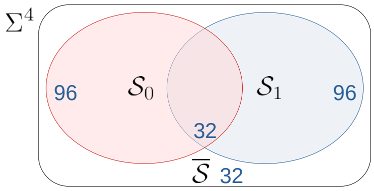
constrained coding

* run length
* GC-balance
* prior: p(x) for IDS

channel coding

* IDS/asymmetric
* multi-read

4-ary



$|\overline{\mathcal{S}}| = 32$

$|\mathcal{S}| = 224$

$|\mathcal{S}_0| = |\mathcal{S}_1| = 128$

$|\mathcal{S}_0 \cap \mathcal{S}_1| = 32$

$\overline{S} = \{\mathsf{A},\mathsf{T}\}^4 \cup \{\mathsf{G},\mathsf{C}\}^4$

$S = S_0 \cup S_1$

$S \cup \overline{S} = \Sigma^4$

0000
0001,0010,0011,0100,0110,0111
0101,1010
1110,1101,1100,1011,1001,1000
1111

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

segment size: β [symbols]

symbol size: $b = 2\beta - 1$ [bits]

inner code rate: $R = (2\beta-1)/2\beta$

set of binary vectors:

$$\mathcal{B}_0 = \{(0, b_1, \dots, b_{\beta-1}) \mid b_i \in \mathbb{B}\} \qquad |\mathcal{B}_0| = |\mathcal{B}_1| = 2^{\beta-1}$$
$$\mathcal{B}_1 = \{(1, b_1, \dots, b_{\beta-1}) \mid b_i \in \mathbb{B}\} \qquad |\mathcal{B}_\mathrm{a}| = |\mathcal{B}_\mathrm{c}| = 2$$
$$\mathcal{B}_\mathrm{a} = \left\{(b, \bar{b}, b, \bar{b}, \dots) \in \mathbb{B}^\beta \,\middle|\, b \in \mathbb{B}\right\}$$
$$\mathcal{B}_\mathrm{c} = \left\{(b, b, b, b, \dots) \in \mathbb{B}^\beta \,\middle|\, b \in \mathbb{B}\right\}$$
$$\tilde{\mathcal{B}}_i = (\mathcal{B}_i \setminus \mathcal{B}_\mathrm{c}) \cup \mathcal{B}_\mathrm{a} \quad (i \in \mathbb{B})$$

segment: $\boldsymbol{s} = (s_0, s_1, \dots, s_{\beta-1}) \in \Sigma^\beta$

symbol mapping: $\phi : \mathbb{B} \times \mathbb{B} \to \Sigma$

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

$$\boldsymbol{\phi} : \mathbb{B}^\beta \times \mathbb{B}^\beta \to \Sigma^\beta$$
$$\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y}) = (\phi(x_0, y_0), \dots, \phi(x_{\beta-1}, y_{\beta-1}))$$
$$\boldsymbol{x} = (x_0, \dots, x_{\beta-1}) \in \mathbb{B}^\beta$$
$$\boldsymbol{y} = (y_0, \dots, y_{\beta-1}) \in \mathbb{B}^\beta$$

set of segments:

$$\overline{\mathcal{S}} = \left\{\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y}) \,\middle|\, \boldsymbol{x} \in \mathcal{B}_\mathrm{c}, \boldsymbol{y} \in \mathbb{B}^\beta\right\} = \{\mathsf{A},\mathsf{T}\}^\beta \cup \{\mathsf{G},\mathsf{C}\}^\beta$$
$$\mathcal{S}_i = \left\{\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y}) \,\middle|\, \boldsymbol{x} \in \tilde{\mathcal{B}}_i, \boldsymbol{y} \in \mathbb{B}^\beta\right\} \quad (i \in \mathbb{B})$$

$(\mathcal{S}_0 \cup \mathcal{S}_1) = (\Sigma^\beta \setminus \overline{\mathcal{S}})$

$|\mathcal{S}_0| = |\mathcal{S}_1| = 2^b$

$|\mathcal{S}_0 \cap \mathcal{S}_1| = 2^{\beta+1}$

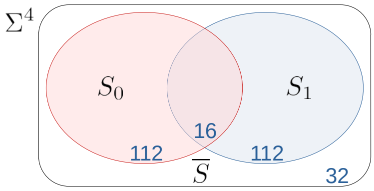
$|\overline{\mathcal{S}}| = 2^{\beta+1}$

encoding func.:

$$\psi_i : \mathbb{B}^{\beta-1} \times \mathbb{B}^\beta \to \mathcal{S}_i \quad (i \in \mathbb{B})$$
$$\psi_i(\tilde{\boldsymbol{x}}, \boldsymbol{y}) = \boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y})$$
$$\boldsymbol{x} = \begin{cases} (i, \tilde{\boldsymbol{x}}) & ((i, \tilde{\boldsymbol{x}}) \notin \mathcal{B}_\mathrm{c}) \\ (\bar{i}, i, \bar{i}, i, \dots) & ((i, \tilde{\boldsymbol{x}}) \in \mathcal{B}_\mathrm{c}) \end{cases}$$

(base caller)
(NB marker)

set partitioning:



$|\overline{S}| = 32$

$|S| = 224$

$|S_0| = |S_1| = 128$

$|S_0 \cap S_1| = 16$

$\overline{S} = \{\mathsf{A}, \mathsf{T}\}^4 \cup \{\mathsf{G}, \mathsf{C}\}^4$

$S = S_0 \cup S_1$

$S \cup \overline{S} = \Sigma^4$

$S_0 \cap S_1 =$

$\left\{ x_0 x_1 x_2 x_3 \middle| x_0 x_2 \in \{\mathsf{A}, \mathsf{T}\}^2, x_1 x_3 \in \{\mathsf{G}, \mathsf{C}\}^2 \right\}$