Parameters

[channel]

$\Sigma = \{\mathsf{A}, \mathsf{T}, \mathsf{G}, \mathsf{C}\}$: channel alphabet
p_{i} : insertion error prob.
p_{d} : deletion error prob.
$p_{\mathrm{s}}(y|x)$: substitution error prob. $(x, y \in \Sigma)$

[constraint]

ℓ_{r} : maximum run-length
$(\ell_{\mathrm{b}}, \varepsilon)$: local GC-balance

$$\frac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_i^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \frac{1}{2} + \varepsilon$$

$$\underline{w} = \left\lceil \ell_b \left(\tfrac{1}{2} - \varepsilon\right) \right\rceil, \ \overline{w} = \left\lfloor \ell_b \left(\tfrac{1}{2} + \varepsilon\right) \right\rfloor$$

$|i| = i \bmod \nu$

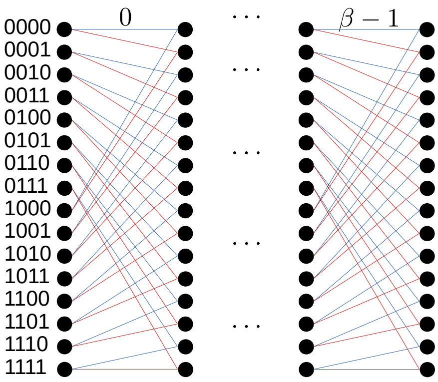
xy	00	01	10	11
$\psi(xy)$	A	T	G	C

Inner codebook

β : code length
ν : number of codebooks
λ : encoder memory $\lambda = \max\{\ell_{\mathrm{r}}, \ell_{\mathrm{b}}\} - 1$
$\mathcal{S}$: set of states $\mathcal{S} \subseteq \mathbb{B}^{\lambda}$
$\boldsymbol{b} = (b_0, \ldots, b_{\nu-1})$: information bit length $(0 \leq b_i \leq 2\beta)$

$\mathcal{C}_{i,s} \subset \Sigma^{\beta}$: codebook $(i \in [\nu\rangle, s \in \mathcal{S})$
$|\mathcal{C}_{i,s}| = 2^{b_i}$

[constraint trellis]

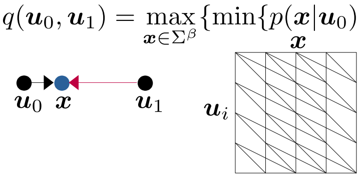


(1) delete nodes/paths of forbidden patterns
$\mathcal{B}_{i,s} \subset \mathbb{B}^{\beta}$: set of valid vectors

(2) codeword search

$$q(\boldsymbol{u}_0, \boldsymbol{u}_1) = \max_{\boldsymbol{x} \in \Sigma^{\beta}} \{\min\{p(\boldsymbol{x}|\boldsymbol{u}_0), p(\boldsymbol{x}|\boldsymbol{u}_1)\}\}$$

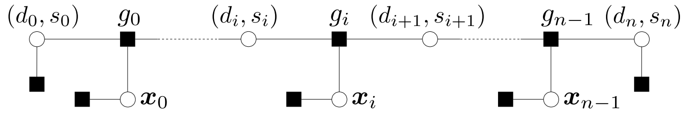
$(\boldsymbol{u}_0 \neq \boldsymbol{u}_1)$



1. initialize:
$$\mathcal{C}_{i,s} = \left\{\boldsymbol{u} \in \Sigma^{\beta} \middle| u_i = \psi(x_i, y_i), \boldsymbol{x} \in \mathcal{B}_{i,s}, \boldsymbol{y} \in \mathbb{B}^{\beta}\right\}$$
2. delete one word from $\mathcal{C}_{i,s}$:
$$\tilde{\boldsymbol{u}} = \arg\max_{\boldsymbol{u} \in \mathcal{C}_{i,s}} \left(\max_{\boldsymbol{u}' \neq \boldsymbol{u}} q(\boldsymbol{u}, \boldsymbol{u}') \right)$$
3. repeat until $|\mathcal{C}_{i,s}| = 2^{b_i}$

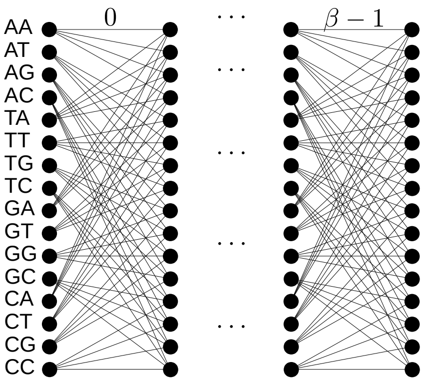
codeword sequence: $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1})$ $\quad \boldsymbol{x}_i \in \mathcal{C}_{|i|, s_i}$
state sequence: $\boldsymbol{s} = (s_0, s_1, \ldots, s_{n-1}, s_n)$ $\quad s_i \in \mathcal{S}$

$$s_i = \begin{cases} c & (i = 0) \text{ (const)} \\ [\boldsymbol{x}_{i-1}]_{\beta-\lambda}^{\beta-1} & (i > 0) \end{cases}$$



$$\begin{aligned} & p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d}) \\ & = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d}) p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d}) \\ & = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{d}) p(\boldsymbol{x}, \boldsymbol{s}) p(\boldsymbol{d}) \\ & = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{d}) p(\boldsymbol{x}|\boldsymbol{s}) p(\boldsymbol{s}) p(\boldsymbol{d}) \\ & = p(s_0) p(d_0) \prod_{i=0}^{n-1} p\left(\boldsymbol{y}_{I_0}^{I_1} \middle| \boldsymbol{x}_i, d_i, d_{i+1}\right) p(\boldsymbol{x}_i|s_i) p(s_{i+1}|s_i) p(d_{i+1}|d_i) \end{aligned}$$

$$\left[\begin{aligned} I_0 &= i\beta + d_i \\ I_1 &= i\beta + \beta - 1 + d_{i+1} \end{aligned} \right.$$



AA
AT
AG
AC
TA
TT
TG
TC
GA
GT
GG
GC
CA
CT
CG
CC
0
$\beta-1$

ICB generation algorithm

xy	00	01	10	11
$\psi(xy)$	A	T	G	C

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

β : block length

ν : number of codebooks $\quad (\beta\nu \bmod 2 = 0)$

δ : max GC-skew $\quad (\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$

$\boldsymbol{b} = (b_0, \ldots, b_{\nu-1})$: information block length $(b_i \geq \beta)$

$\boldsymbol{b}' = (b'_0, \ldots, b'_{\nu-1})$: $b'_i = b_i - \beta$

$\mathcal{B}_i \subseteq \mathbb{B}^\beta$: binary code $|\mathcal{B}_i| = 2^{b'_i} \quad (i \in [\nu\rangle)$

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

$|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

Full search of $(\mathcal{B}_0, \ldots, \mathcal{B}_{\nu-1})$

(1) initialize: $\mathcal{B}_i = \mathbb{B}^\beta$

(2) GC-weight constraint: $\forall \boldsymbol{u} \in \mathcal{B}_i$

$$w(\boldsymbol{u}) = \begin{cases} \lfloor \beta/2 \rfloor & (i \bmod 2 = 0) \\ \lceil \beta/2 \rceil & (i \bmod 2 = 1) \end{cases}$$

(3) GC-skew constraint: $\forall \boldsymbol{u} \in \mathcal{C}_i$

$$|w(\boldsymbol{u}_\mathrm{L}) - w(\boldsymbol{u}_\mathrm{R})| \leq \delta$$
$$\boldsymbol{u}_L = \boldsymbol{u}_0^{\lfloor \beta/2 \rfloor - 1}, \boldsymbol{u}_R = \boldsymbol{u}_{\beta - \lfloor \beta/2 \rfloor}^{\beta-1}$$

(x) additional constraint (?)

(4) generation fails if $|\mathcal{B}_i| < 2^{b'_i}$

(5) full search: maximize R_hd

select from $\prod_{i=0}^{\nu-1} \binom{|\mathcal{B}_i|}{2^{b'_i}}$ patterns

(*) motif: eliminated by mask (=QR code)

Local GC-balance

$(\ell_\mathrm{b}, \varepsilon)$ constraint:

$$\frac{1}{2} - \varepsilon \leq \frac{w_\mathrm{GC}(\boldsymbol{u}_i^{i+\ell_\mathrm{b}-1})}{\ell_\mathrm{b}} \leq \frac{1}{2} + \varepsilon$$
$$\underline{w} = \left\lceil \ell_b \left(\tfrac{1}{2} - \varepsilon\right) \right\rceil$$
$$\overline{w} = \left\lfloor \ell_b \left(\tfrac{1}{2} + \varepsilon\right) \right\rfloor$$