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
$(\beta\nu \bmod 2 = 0)$
$\boldsymbol{b}=(b_0,\dots,b_{\nu-1})$: information block length $(b_i \geq \beta)$
$\mathcal{B}_i \subseteq \mathbb{B}^\beta$: binary code
$\mathcal{C}_i \subseteq \left\{(\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1}))\middle|\boldsymbol{x}\in\mathcal{B}_i,\boldsymbol{y}\in\mathbb{B}^\beta\right\}$
$|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

Greedy search of $(\mathcal{C}_0,\dots,\mathcal{C}_{\nu-1})$

(1) intra-word constraint: determine $\mathcal{B}_i \subset \mathbb{B}^\beta$ satisfying

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

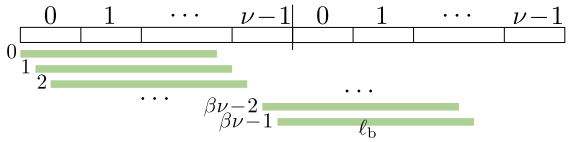
(ii) $f_{\mathrm{r}}(\boldsymbol{u}) \leq \ell_{\mathrm{r}}$ (run-length)

(iii) $u_0=0, u_{\beta-1}=1$ (sync + inter RL)

for all $\boldsymbol{u}\in\mathcal{B}_i$

(2) inter-word L-GCB constraint:

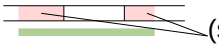
(a) calculate weights: $w_{\max}(i), w_{\min}(i)$ $(i\in[\beta\nu\rangle)$



(b) if $\forall i\in[\beta\nu\rangle, w_{\min}(i)\geq\underline{w}, w_{\max}(i)\leq\overline{w}$, go to (3).

(c) select worst position:
$\tilde{i} = \underset{i\in[\beta\nu\rangle}{\arg\max}\{|w_{\max}(i)-\frac{\ell_{\mathrm{b}}}{2}|, |w_{\min}(i)-\frac{\ell_{\mathrm{b}}}{2}|\}$

(d) delete worst words:


(select larger skew)

(e) go to (a)

(3) initial ICB:
$\mathcal{C}_i = \left\{(\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1}))\middle|\boldsymbol{x}\in\mathcal{B}_i,\boldsymbol{y}\in\mathbb{B}^\beta\right\}$

(4) remove motif: [?]
$\mathcal{C}_i = \mathcal{C}_i \setminus \mathcal{X}_i$

(5) generation failed if $\exists i\in[\nu\rangle, |\mathcal{C}_i| < 2^{b_i}$

(6) remove large metric codewords:

$$\tilde{q}(\boldsymbol{y}|\boldsymbol{u}) = \begin{cases}0 & (p(\boldsymbol{y}|\boldsymbol{u}) = \max_{\tilde{\boldsymbol{u}}\in\mathcal{C}_i} p(\boldsymbol{y}|\tilde{\boldsymbol{u}}))\\ p(\boldsymbol{y}|\boldsymbol{u}) & (\text{otherwise})\end{cases}$$
$$(\boldsymbol{u}\in\mathcal{C}_i, \boldsymbol{y}\in\Sigma^\beta)$$

$$\mathcal{Y}_i = \underset{\boldsymbol{u}\in\mathcal{C}_i}{\arg\max\,\mathrm{list}_t}\left\{\max_{\boldsymbol{y}\in\Sigma^\beta}\tilde{q}(\boldsymbol{y}|\boldsymbol{u})\right\} \quad (t = |\mathcal{C}_i| - 2^{b_i})$$

$\mathcal{C}_i = \mathcal{C}_i \setminus \mathcal{Y}_i$

Local GC-balance

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

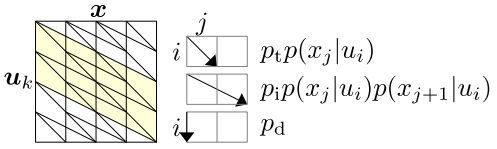
$$\tfrac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_i^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \tfrac{1}{2} + \varepsilon$$

$\underline{w} = \lceil \ell_b\left(\frac{1}{2}-\varepsilon\right)\rceil$

$\overline{w} = \lfloor \ell_b\left(\frac{1}{2}+\varepsilon\right)\rfloor$

Metric

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



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
$(\beta\nu \bmod 2 = 0)$
$\boldsymbol{b}=(b_0,\dots,b_{\nu-1})$: information block length $(b_i \geq \beta)$
$\mathcal{B}_i \subseteq \mathbb{B}^\beta$: binary code
$\mathcal{C}_i \subseteq \left\{(\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1}))\middle|\boldsymbol{x}\in\mathcal{B}_i,\boldsymbol{y}\in\mathbb{B}^\beta\right\}$
$|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

Greedy search of $(\mathcal{C}_0,\dots,\mathcal{C}_{\nu-1})$

(1) intra-word constraint: determine $\mathcal{B}_i \subset \mathbb{B}^\beta$ satisfying

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

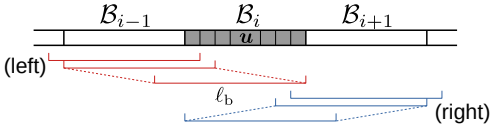
(ii) $f_{\mathrm{r}}(\boldsymbol{u}) \leq \ell_{\mathrm{r}}$ (run-length)

(iii) $u_0=0, u_{\beta-1}=1$ (sync + inter RL)

for all $\boldsymbol{u}\in\mathcal{B}_i$

(2) inter-word L-GCB constraint:

(a) delete worst score codeword $\boldsymbol{u}\in\mathcal{B}_i$



score = # of unsatisfied patterns (left+right)
tie-break: $|\mathcal{B}_i| - 2^{b_i}$

(b) $\mathcal{B}_i = \mathcal{B}_i \setminus \{\boldsymbol{u}\}$

(c) repeat until the constraint is satisfied

(3) initial ICB:
$\mathcal{C}_i = \left\{(\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1}))\middle|\boldsymbol{x}\in\mathcal{B}_i,\boldsymbol{y}\in\mathbb{B}^\beta\right\}$

(4) remove motif: [?]
$\mathcal{C}_i = \mathcal{C}_i \setminus \mathcal{X}_i$

(5) generation failed if $\exists i \in [\nu\rangle, |\mathcal{C}_i| < 2^{b_i}$

(6) remove large metric codewords:

$$\tilde{q}(\boldsymbol{y}|\boldsymbol{u}) = \begin{cases}0 & (p(\boldsymbol{y}|\boldsymbol{u}) = \max_{\tilde{\boldsymbol{u}}\in\mathcal{C}_i} p(\boldsymbol{y}|\tilde{\boldsymbol{u}}))\\ p(\boldsymbol{y}|\boldsymbol{u}) & (\text{otherwise})\end{cases} \quad (\boldsymbol{u}\in\mathcal{C}_i, \boldsymbol{y}\in\Sigma^\beta)$$

$$\mathcal{Y}_i = \arg\max_{\boldsymbol{u}\in\mathcal{C}_i} \mathrm{list}_t \left\{\max_{\boldsymbol{y}\in\Sigma^\beta} \tilde{q}(\boldsymbol{y}|\boldsymbol{u})\right\} \quad (t = |\mathcal{C}_i| - 2^{b_i})$$

$\mathcal{C}_i = \mathcal{C}_i \setminus \mathcal{Y}_i$

Local GC-balance

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

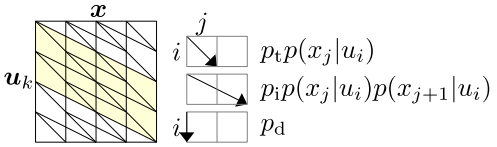
$$\tfrac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_i^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \tfrac{1}{2} + \varepsilon$$

$$\underline{w} = \left\lceil \ell_b \left(\tfrac{1}{2} - \varepsilon\right)\right\rceil$$

$$\overline{w} = \left\lfloor \ell_b \left(\tfrac{1}{2} + \varepsilon\right)\right\rfloor$$

Metric

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



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
$(\beta\nu \bmod 2 = 0)$
$\boldsymbol{b}=(b_0,\dots,b_{\nu-1})$: information block length $(b_i \geq \beta)$
$\mathcal{B}_i \subseteq \mathbb{B}^\beta$: binary code
$\mathcal{C}_i \subseteq \left\{(\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1}))\middle|\boldsymbol{x}\in\mathcal{B}_i,\boldsymbol{y}\in\mathbb{B}^\beta\right\}$
$|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

Greedy search of $(\mathcal{C}_0,\dots,\mathcal{C}_{\nu-1})$

(1) intra-word constraint: determine $\mathcal{B}_i \subset \mathbb{B}^\beta$ satisfying

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

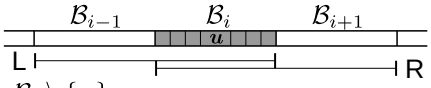
(ii) $f_{\mathrm{r}}(\boldsymbol{u}) \leq \ell_{\mathrm{r}}$ (run-length)

(iii) $u_0=0, u_{\beta-1}=1$ (sync)

for all $\boldsymbol{u}\in\mathcal{B}_i$

(3) inter-word RL constraint:

(a) delete worst score codeword $\boldsymbol{u}\in\mathcal{B}_i$

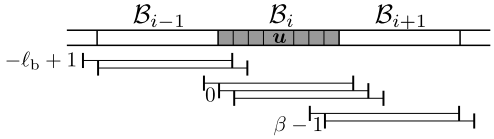


(b) $\mathcal{B}_i = \mathcal{B}_i \setminus \{\boldsymbol{u}\}$

(c) repeat until the constraint is satisfied

(2) inter-word LGCB constraint:

(a) delete worst score codeword $\boldsymbol{u}\in\mathcal{B}_i$



(b) $\mathcal{B}_i = \mathcal{B}_i \setminus \{\boldsymbol{u}\}$

(c) repeat until the constraint is satisfied

(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:

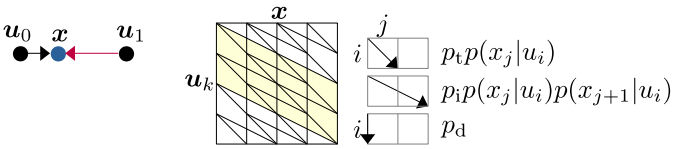
$\frac{1}{2}-\varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_i^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \frac{1}{2}+\varepsilon$

$\underline{w} = \lceil \ell_b\left(\frac{1}{2}-\varepsilon\right)\rceil$

$\overline{w} = \lfloor \ell_b\left(\frac{1}{2}+\varepsilon\right)\rfloor$

Metric

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



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
δ : max GC-skew

$(\beta\nu \bmod 2 = 0)$
$(\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$

$\boldsymbol{b} = (b_0, \ldots, b_{\nu-1})$: 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 $|\mathcal{B}_i| = 2^{b'_i} \quad (i \in [\nu\rangle)$
$\mathcal{C}_i = \left\{ (\psi(x_0y_0), \ldots, \psi(x_{\beta-1}y_{\beta-1})) \middle| \boldsymbol{x} \in \mathcal{B}_i, \boldsymbol{y} \in \mathbb{B}^\beta \right\}$
$|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: 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: $\forall \boldsymbol{u} \in \mathcal{C}_i$

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

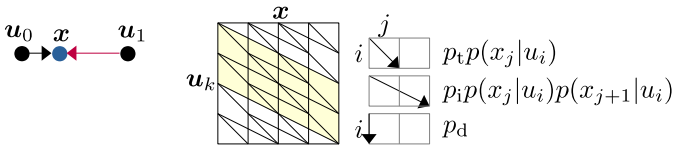
(x) additional constraint (?)
(4) generation fails if $|\mathcal{B}_i| < 2^{b'_i}$
(5) full search: maximize R_hd

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

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

Metric

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



Local GC-balance

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

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