ICB generation algorithm

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

$$\beta$$
: block length

$$\nu$$
 : number of codebooks

$$(\beta\nu \bmod 2 = 0)$$

$$\delta: \max \mathsf{GC}\text{-skew}$$

$$(\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$$

$$b = (b_0, \dots, b_{\nu-1}) : info$$

$$(\delta \in \lfloor \lfloor \beta/2 \rfloor + 1)$$

$$\mathcal{C} \subset \Sigma^{\beta}$$
 : codobook

$$m{b} = (b_0, \dots, b_{\nu-1})$$
: information block length $\mathcal{C}_i \subseteq \Sigma^\beta$: codebook $|\mathcal{C}_i| = 2^{b_i} \le 4^\beta$: number of codewords

(1) initialize:
$$C_i := \Sigma^{\beta}$$

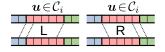
(1) initialize:
$$\mathcal{C}_i := \Sigma^{\beta}$$
 (2) GC-weight constraint: $\forall \boldsymbol{u} \in \mathcal{C}_i$

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

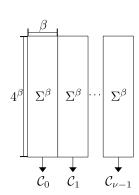
(3) GC-skew constraint: $\forall u \in \mathcal{C}_i$

$$\begin{aligned} |w_{\mathsf{GC}}(\boldsymbol{u}_{\mathrm{L}}) - w_{\mathsf{GC}}(\boldsymbol{u}_{\mathrm{R}})| &\leq \delta \\ \boldsymbol{u}_{L} &= \boldsymbol{u}_{0}^{\lfloor \beta/2 \rfloor - 1}, \boldsymbol{u}_{R} = \boldsymbol{u}_{\beta - \lfloor \beta/2 \rfloor}^{\beta - 1} \end{aligned}$$

- (x) additional constraint (?)
- (4) generation fails if $|\mathcal{C}_i| < 2^{b_i}$
- (5) select 2^{b_i} vectors from C_i



- 1) count distance-0 patterns
- 2) delete max
- (*) motif: eliminated by mask (=QR code)



Local GC-balance

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

$$\frac{1}{2} - \varepsilon \leq \frac{w_{\text{GC}}(u_i^{i+\ell_{\text{b}}-1})}{\ell_{\text{b}}} \leq \frac{1}{2} + \varepsilon$$

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

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

ICB generation algorithm

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

 β : block length

 ν : number of codebooks

 $(\beta \nu \mod 2 = 0)$

 $\delta: \max \mathsf{GC}\text{-skew}$

 $(\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$

 $m{b} = (b_0, \dots, b_{\nu-1})$: information block length $\mathcal{C}_i \subseteq \Sigma^\beta$: codebook $|\mathcal{C}_i| = 2^{b_i} \le 4^\beta$: number of codewords

(1) initialize: $\mathcal{C}_i:=\Sigma^{\beta}$ (2) GC-weight constraint: $\forall m{u}\in\mathcal{C}_i$

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

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

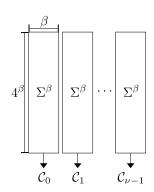
$$egin{align*} |w_{\mathsf{GC}}(oldsymbol{u}_L) - w_{\mathsf{GC}}(oldsymbol{u}_R)| \leq \delta \ oldsymbol{u}_L = oldsymbol{u}_0^{\lfloor eta/2
floor}, oldsymbol{u}_R = oldsymbol{u}_{eta-\lfloor eta/2
floor}^{eta-1}, \end{align}$$

- (x) additional constraint (?)
- (4) generation fails if $|\mathcal{C}_i| < 2^{b_i}$
- (5) select 2^{b_i} vectors from C_i

lower symbol-wise entropy (greedy?)

$$\sum_{j=0}^{\beta-1} H(U_j)$$
 (tie-break by RNG?)

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



Local GC-balance

$$\begin{array}{l} (\ell_{\mathrm{b}},\varepsilon) \text{ constraint:} \\ \frac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_{i}^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \frac{1}{2} + \varepsilon \end{array}$$

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

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

ICB generation algorithm

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

 β : block length

 ν : number of codebooks

 $(\beta \nu \mod 2 = 0)$

 $\delta: \max \mathsf{GC}\text{-skew}$

 $(\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$

 $F_{i,j} \subseteq \Sigma$: forbidden symbol set $(i \in [\nu\rangle, j \in [\beta\rangle)$ $\mathcal{C}_i \subseteq \Sigma^{\beta}$: codebook $|\mathcal{C}_i| = 2^{b_i} \le 4^{\beta}$: number of codewords

 b_i : information block length $(i \in [\nu\rangle)$

(1) initialize: $C_i := \Sigma^{\beta}$

(2) GC-weight constraint: $orall oldsymbol{u} \in \mathcal{C}_i$

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

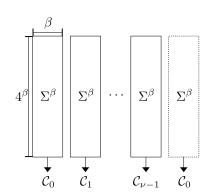
(3) GC-skew constraint: $\forall u \in \mathcal{C}_i$

$$egin{align*} |w_{\mathsf{GC}}(oldsymbol{u}_L) - w_{\mathsf{GC}}(oldsymbol{u}_R)| \leq \delta \ oldsymbol{u}_L = oldsymbol{u}_0^{\lfloor eta/2
floor}, oldsymbol{u}_R = oldsymbol{u}_{eta-\lfloor eta/2
floor}^{eta-1}, \end{align}$$

(4) forbidden symbol:

$$\forall \mathbf{u} = (u_0, \dots, u_{\beta-1}) \in \mathcal{C}_i, u_j \notin F_{i,j}$$

- (5) information length: $b_i = |\log_2 |\mathcal{C}_i||$
- (6) select 2^{b_i} vectors from C_i
- (*) motif: eliminated by mask (=QR code)



Local GC-balance

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

$$\begin{split} &(\ell_{\mathrm{b}},\varepsilon) \text{ constraint:} \\ &\frac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_{i}^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \frac{1}{2} + \varepsilon \\ &\underline{w} = \left\lceil \ell_{b} \left(\frac{1}{2} - \varepsilon\right) \right\rceil \\ &\overline{w} = \left\lvert \ell_{b} \left(\frac{1}{2} + \varepsilon\right) \right\rvert \end{split}$$

(num. of GC-balanced patterns)
$$|\mathcal{C}_i| = \begin{cases} 2^{\beta} \binom{\beta}{(\beta-1)/2} & (\text{odd } \beta) \\ 2^{\beta} \binom{\beta}{\beta/2} & (\text{even } \beta) \end{cases}$$

ICB search algorithm

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

 β : block length

 ν : number of codebooks

 b_i : information block length $(i \in [
u\rangle)$

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

- (1) initialize: $C_i := \Sigma^{\beta}$
- (2) remove GC-imbalance words (intra-word):
- (3) remove RL-violating words (intra-word)
- (4) remove constraint violating words (inter-word)

* local GCB

* RL

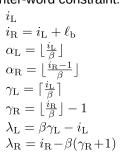
(score base?)

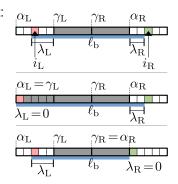
- (5) select marker-like subset (for synchronization) [marker + anti-marker?]
- (*) motif: eliminated by mask (QR code)

Local GC-balance

$$\begin{split} (\ell_{\mathrm{b}}, \varepsilon) \text{ constraint: } & \frac{1}{2} - \varepsilon \leq \frac{w_{\mathrm{GC}}(\boldsymbol{u}_{i}^{i+\ell_{\mathrm{b}}-1})}{\ell_{\mathrm{b}}} \leq \frac{1}{2} + \varepsilon \\ & \underline{w} = \left\lceil \ell_{b} \left(\frac{1}{2} - \varepsilon \right) \right\rceil, \overline{w} = \left\lfloor \ell_{b} \left(\frac{1}{2} + \varepsilon \right) \right\rfloor \end{split}$$

inter-word constraint:





(1) window position: $i_L \in [\nu\beta\rangle$

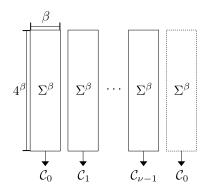
 $i_{\rm R},\alpha_{\rm L},\alpha_{\rm R},\gamma_{\rm L},\gamma_{\rm R},\lambda_{\rm L},\lambda_{\rm R}\ (\alpha_{\rm L}<\alpha_{\rm R})$

(2) center blocks:

 $L_{\rm c} = \beta(\alpha_{\rm R} - \alpha_{\rm L} - 1)$ $W_{\rm c} = \sum_{i=\alpha_{\rm L}+1}^{\alpha_{\rm R}-1} w(\mathcal{C}_{|i|})$

(3) partial codebooks: $\mathcal{C}_{\mathrm{L}} = \left\{m{u}_{eta-\lambda_{\mathrm{L}}}^{eta-1} \middle| m{u} \in \mathcal{C}_{|lpha_{\mathrm{L}}|}
ight\}$

 $\mathcal{C}_{\mathrm{R}} = \left\{oldsymbol{u}_{0}^{\lambda_{\mathrm{R}}-1} \middle| oldsymbol{u} \in \mathcal{C}_{|lpha_{\mathrm{R}}|}
ight\}$



(num. of GC-balanced patterns)

$$|\mathcal{C}_i| = \begin{cases} 2^{\beta} \binom{\beta}{(\beta - 1)/2} & \text{(odd } \beta) \\ 2^{\beta} \binom{\beta}{\beta/2} & \text{(even } \beta) \end{cases}$$

(1) set window position: $oldsymbol{v} = oldsymbol{u}_{i_1}^{i_{
m L}+\ell_{
m b}-1} \in \Sigma^{\ell_{
m b}}$ set balance sign:

$$\sigma = \begin{cases} 0 & (w_{\min} \leq w_{\mathrm{GC}}(\boldsymbol{v}) \leq w_{\max}) & \text{(nop)} \\ -1 & (w_{\mathrm{GC}}(\boldsymbol{v}) < w_{\min}) & \text{(inc)} \\ +1 & (w_{\mathrm{GC}}(\boldsymbol{v}) > w_{\max}) & \text{(dec)} \end{cases}$$

(2)