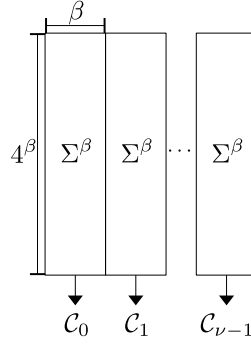


ICB generation algorithm

$\Sigma = \{A, T, G, C\}$: alphabet
 β : block length
 ν : number of codebooks
 δ : max GC-skew
 $\mathbf{b} = (b_0, \dots, b_{\nu-1})$: information block length
 $\mathcal{C}_i \subseteq \Sigma^\beta$: codebook
 $|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

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

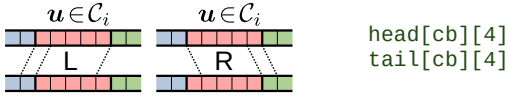


(1) initialize: $\mathcal{C}_i := \Sigma^\beta$
 (2) GC-weight constraint: $\forall \mathbf{u} \in \mathcal{C}_i$

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

(3) GC-skew constraint: $\forall \mathbf{u} \in \mathcal{C}_i$
 $|w_{GC}(\mathbf{u}_L) - w_{GC}(\mathbf{u}_R)| \leq \delta$
 $\mathbf{u}_L = \mathbf{u}_0^{\lfloor \beta/2 \rfloor - 1}, \mathbf{u}_R = \mathbf{u}_{\beta - \lfloor \beta/2 \rfloor}^{\beta - 1}$

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



1) count distance-0 patterns
 2) delete max

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

Local GC-balance

(ℓ_b, ε) constraint:

$$\frac{1}{2} - \varepsilon \leq \frac{w_{GC}(\mathbf{u}_i^{i+\ell_b-1})}{\ell_b} \leq \frac{1}{2} + \varepsilon$$

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

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

ICB generation algorithm

$\Sigma = \{A, T, G, C\}$: alphabet
 β : block length
 ν : number of codebooks
 δ : max GC-skew
 $\mathbf{b} = (b_0, \dots, b_{\nu-1})$: information block length
 $\mathcal{C}_i \subseteq \Sigma^\beta$: codebook
 $|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords

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

(1) initialize: $\mathcal{C}_i := \Sigma^\beta$

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

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

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

$$|w_{GC}(\mathbf{u}_L) - w_{GC}(\mathbf{u}_R)| \leq \delta$$

$$\mathbf{u}_L = \mathbf{u}_0^{\lfloor \beta/2 \rfloor - 1}, \mathbf{u}_R = \mathbf{u}_{\beta - \lfloor \beta/2 \rfloor}^{\beta - 1}$$

(x) additional constraint (?)

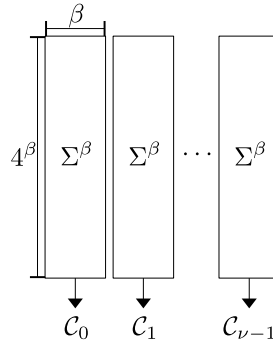
(4) generation fails if $|\mathcal{C}_i| < 2^{b_i}$

(5) select 2^{b_i} vectors from \mathcal{C}_i

lower symbol-wise entropy (greedy?)

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

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



Local GC-balance

(ℓ_b, ε) constraint:

$$\frac{1}{2} - \varepsilon \leq \frac{w_{GC}(\mathbf{u}_i^{i+\ell_b-1})}{\ell_b} \leq \frac{1}{2} + \varepsilon$$

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

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

ICB generation algorithm

$\Sigma = \{A, T, G, C\}$: alphabet
 β : block length
 ν : number of codebooks $(\beta\nu \bmod 2 = 0)$
 δ : max GC-skew $(\delta \in [\lfloor \beta/2 \rfloor + 1])$
 $F_{i,j} \subsetneq \Sigma$: forbidden symbol set $(i \in [\nu], j \in [\beta])$
 $\mathcal{C}_i \subseteq \Sigma^\beta$: codebook
 $|\mathcal{C}_i| = 2^{b_i} \leq 4^\beta$: number of codewords
 b_i : information block length $(i \in [\nu])$

(1) initialize: $\mathcal{C}_i := \Sigma^\beta$

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

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

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

$$|w_{GC}(\mathbf{u}_L) - w_{GC}(\mathbf{u}_R)| \leq \delta$$

$$\mathbf{u}_L = \mathbf{u}_0^{\lfloor \beta/2 \rfloor - 1}, \mathbf{u}_R = \mathbf{u}_{\beta - \lfloor \beta/2 \rfloor}^{\beta - 1}$$

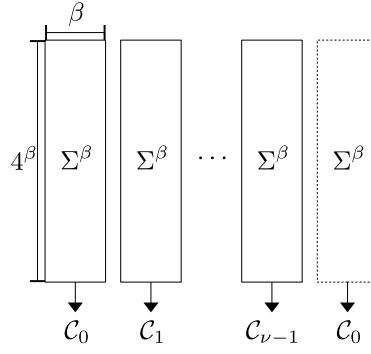
(4) forbidden symbol:

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

(5) information length: $b_i = \lfloor \log_2 |\mathcal{C}_i| \rfloor$

(6) select 2^{b_i} vectors from \mathcal{C}_i

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



Local GC-balance

(ℓ_b, ε) constraint:

$$\frac{1}{2} - \varepsilon \leq \frac{w_{GC}(\mathbf{u}_i^{i+\ell_b-1})}{\ell_b} \leq \frac{1}{2} + \varepsilon$$

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

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

(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

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

β : block length

ν : number of codebooks

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

$\mathcal{C}_i \subseteq \Sigma^\beta$: codebook

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

(1) initialize: $\mathcal{C}_i := \Sigma^\beta$

(2) remove GC-imbalance words (intra-word):

(3) remove RL-violating words (intra-word)

(4) remove constraint violating words (inter-word)

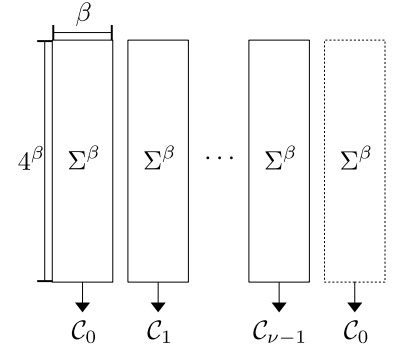
* local GCB (score base?)

* RL

(5) select marker-like subset (for synchronization)

[marker + anti-marker?]

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



(num. of GC-balanced patterns)

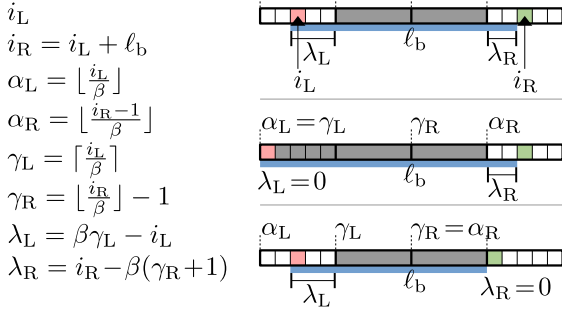
$$|\mathcal{C}_i| = \begin{cases} 2^\beta \binom{\beta}{(\beta-1)/2} & (\text{odd } \beta) \\ 2^\beta \binom{\beta}{\beta/2} & (\text{even } \beta) \end{cases}$$

Local GC-balance

$$(\ell_b, \varepsilon) \text{ constraint: } \frac{1}{2} - \varepsilon \leq \frac{w_{GC}(\mathbf{u}_{i_L + \ell_b - 1})}{\ell_b} \leq \frac{1}{2} + \varepsilon$$

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

inter-word constraint:



(1) window position: $i_L \in [\nu\beta]$
 $i_R, \alpha_L, \alpha_R, \gamma_L, \gamma_R, \lambda_L, \lambda_R$ ($\alpha_L < \alpha_R$)

(2) center blocks: $L_c = \beta(\alpha_R - \alpha_L - 1)$

$$W_c = \sum_{i=\alpha_L+1}^{\alpha_R-1} w(\mathcal{C}_i)$$

(3) partial codebooks: $\mathcal{C}_L = \left\{ \mathbf{u}_{\beta-\lambda_L}^{\beta-1} \mid \mathbf{u} \in \mathcal{C}_{|\alpha_L|} \right\}$
 $\mathcal{C}_R = \left\{ \mathbf{u}_0^{\lambda_R-1} \mid \mathbf{u} \in \mathcal{C}_{|\alpha_R|} \right\}$

(1) set window position: $\mathbf{v} = \mathbf{u}_{i_L + \ell_b - 1}^{i_L + \ell_b - 1} \in \Sigma^{\ell_b}$
 set balance sign:

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

(2)