$$\Sigma = \{\mathsf{A},\mathsf{T},\mathsf{G},\mathsf{C}\} : \mathsf{alphabet} \\ \beta : \mathsf{block} \ \mathsf{length} \\ \nu : \mathsf{number} \ \mathsf{of} \ \mathsf{codebooks} \\ b = (b_0,\dots,b_{\nu-1}) : \mathsf{information} \ \mathsf{block} \ \mathsf{length} \ (b_i \geq \beta) \\ \mathcal{B}_i \subseteq \mathbb{B}^\beta : \mathsf{binary} \ \mathsf{code} \\ \mathcal{C}_i \subseteq \big\{ (\psi(x_0y_0),\dots,\psi(x_{\beta-1}y_{\beta-1})) \big| \boldsymbol{x} \in \mathcal{B}_i, \boldsymbol{y} \in \mathbb{B}^\beta \big\} \\ |\mathcal{C}_i| = 2^{b_i} \leq 4^\beta : \mathsf{number} \ \mathsf{of} \ \mathsf{codewords} \\ \hline$$

Greedy search of $(C_0, \ldots, C_{\nu-1})$

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

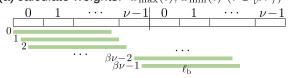
(i)
$$w(u) = \begin{cases} \lfloor \beta/2 \rfloor & (i \bmod 2 = 0) \\ \lceil \beta/2 \rceil & (i \bmod 2 = 1) \end{cases}$$
 (GC-balance)

- (run-length)

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

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

- (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:

$$\tilde{i} = \operatorname*{arg\,max}_{i \in [\beta \nu \rangle} \{ |w_{\max}(i) - \tfrac{\ell_{\mathrm{b}}}{2}|, |w_{\min}(i) - \tfrac{\ell_{\mathrm{b}}}{2}| \}$$

(d) delete worst words:

- (e) go to (a)
- (3) initial ICB:

$$\mathcal{C}_i = \left\{ (\psi(x_0 y_0), \dots, \psi(x_{eta-1} y_{eta-1})) \middle| oldsymbol{x} \in \mathcal{B}_i, oldsymbol{y} \in \mathbb{B}^{eta}
ight\}$$

(4) remove motif: [?]

$$C_i = C_i \setminus \mathcal{X}_i$$

- (5) generation failed if $\exists i \in [\nu\rangle, |\mathcal{C}_i| < 2^{b_i}$
- (6) remove large metric codewords:

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

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

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

Local GC-balance

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

$$\frac{1}{2} - \varepsilon \le \frac{w_{\text{GC}}(\boldsymbol{u}_{i}^{i+\ell_{\text{b}}-1})}{\ell_{\text{b}}} \le \frac{1}{2} + \varepsilon$$

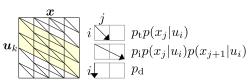
$$\underline{w} = \lceil \ell_{b} \left(\frac{1}{2} - \varepsilon \right) \rceil$$

$$\overline{w} = \left| \ell_{b} \left(\frac{1}{2} + \varepsilon \right) \right|$$

Metric

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

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



$$\Sigma = \{\mathsf{A},\mathsf{T},\mathsf{G},\mathsf{C}\} : \mathsf{alphabet} \qquad \frac{xy \quad 00 \quad 01 \quad 10 \quad 11}{\psi(xy) \quad \mathsf{A} \quad \mathsf{T} \quad \mathsf{G} \quad \mathsf{C}}$$

$$\beta : \mathsf{block} \; \mathsf{length} \qquad \qquad (\beta \nu \bmod 2 = 0)$$

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

$$\mathcal{B}_i \subseteq \mathbb{B}^\beta : \; \mathsf{binary} \; \mathsf{code}$$

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

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

Greedy search of $(C_0, \ldots, C_{\nu-1})$

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

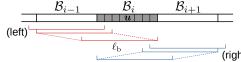
$$\begin{array}{ll} \text{(i)} \quad w(\boldsymbol{u}) = \begin{cases} \lfloor \beta/2 \rfloor & (i \bmod 2 = 0) \\ \lceil \beta/2 \rceil & (i \bmod 2 = 1) \end{cases} \text{ (GC-balance)} \\ \text{(ii)} \quad f_{\mathrm{r}}(\boldsymbol{u}) \leq \ell_{\mathrm{r}} \qquad \qquad \text{(run-length)} \end{array}$$

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

(sync + inter RL)

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

(2) inter-word L-GCB constraint: (a) delete worst score codeword $u \in \mathcal{B}_i$



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

- **(b)** $\mathcal{B}_i = \mathcal{B}_i \setminus \{ oldsymbol{u} \}$
- (c) repeat until the constraint is satisfied
- (3) initial ICB:

$$C_i = \{ (\psi(x_0 y_0), \dots, \psi(x_{\beta-1} y_{\beta-1})) | \boldsymbol{x} \in \mathcal{B}_i, \boldsymbol{y} \in \mathbb{B}^{\beta} \}$$

(4) remove motif: [?]

$$C_i = C_i \setminus \mathcal{X}_i$$

- (5) generation failed if $\exists i \in [\nu\rangle, |\mathcal{C}_i| < 2^{b_i}$
- (6) remove large metric codewords:

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

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

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

Local GC-balance

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

$$\frac{1}{2} - \varepsilon \le \frac{w_{\text{GC}}(\boldsymbol{u}_{i}^{i+\ell_{\text{b}}-1})}{\ell_{\text{b}}} \le \frac{1}{2} + \varepsilon$$

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

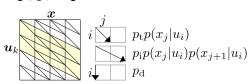
$$\overline{w} = \left| \ell_{b} \left(\frac{1}{2} + \varepsilon \right) \right|$$

Metric

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

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

$$\boldsymbol{x} \quad j$$



CR generation algorithm								
CB generation algorithm	xy	00	01	10	11			
$\Sigma = \{A, T, G, C\}$: alphabet	$\psi(xy)$	Α	Т	G	С			
β : block length	mod S	_	0)					
ν : number of codebooks)	$(\beta\nu \bmod 2 = 0)$							
$b = (b_0, \dots, b_{\nu-1})$: information block length $(b_i \ge \beta)$								
$\mathcal{B}_i\subseteq\mathbb{B}^eta$: binary code								
$\mathcal{C}_i \subseteq \left\{ (\psi(x_0 y_0), \dots, \psi(x_{\beta-1} y_{\beta-1})) \middle oldsymbol{x} \in \mathcal{B}_i, oldsymbol{y} \in \mathbb{B}^{eta} \right\}$								
$ \mathcal{C}_i = 2^{b_i} \leq 4^{eta}$: number of codewords								

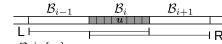
Greedy search of $(\mathcal{C}_0,\ldots,\mathcal{C}_{\nu-1})$

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

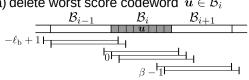
$$\begin{aligned} & \text{(i)} \quad w(\boldsymbol{u}) = \begin{cases} \lfloor \beta/2 \rfloor & (i \bmod 2 = 0) \\ \lceil \beta/2 \rceil & (i \bmod 2 = 1) \end{cases} & \text{(GC-balance)} \\ & \text{(ii)} \quad f_{\mathbf{r}}(\boldsymbol{u}) \leq \ell_{\mathbf{r}} & \text{(run-length)} \\ & \text{(iii)} \quad u_0 = 0, u_{\beta-1} = 1 & \text{(sync)} \end{aligned}$$

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

- (3) inter-word RL constraint:
 - (a) delete worst score codeword $\, oldsymbol{u} \in \mathcal{B}_i \,$



- **(b)** $\mathcal{B}_i = \mathcal{B}_i \setminus \{ oldsymbol{u} \}$
- (c) repeat until the constraint is satisfied
- (2) inter-word LGCB constraint:
 - (a) delete worst score codeword $\, oldsymbol{u} \in \mathcal{B}_i \,$



- **(b)** $\mathcal{B}_i = \mathcal{B}_i \setminus \{ \boldsymbol{u} \}$
- (c) repeat until the constraint is satisfied

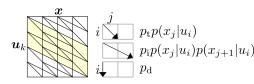
Local GC-balance

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

Metric

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





- (x) additional constraint (?)
- (4) generation fails if $|\mathcal{B}_i| < 2^{b_i'}$
- (5) full search: maximize $R_{\rm hd}$

select from
$$\prod_{i=0}^{
u-1} {|\mathcal{B}_i| \choose 2^{b_i'}}$$
 patterns

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

CB generation algorithm	\overline{xy}	00	01	10	11	
$\Sigma = \{A, T, G, C\}$: alphabet	$\psi(xy)$	Α	Т	G	С	
β : block length	$(\beta \nu \text{ mo})$	od 2		0)		
$ u$: number of codebooks \parallel	C)					
	$(\delta \in [\lfloor \beta/2 \rfloor + 1\rangle)$					
$\boldsymbol{b} = (b_0, \dots, b_{\nu-1})$: information block length $(b_i \geq \beta)$						
$b' = (b'_0, \dots, b'_{\nu-1}) : b'_i = b_i - \beta$						
${\mathcal B}_i\subseteq {\mathbb B}^eta$: binary code $ {\mathcal B}_i =2^{b_i'}$ $(i\in [u angle)$						
$C_i = \{ (\psi(x_0 y_0), \dots, \psi(x_{\beta-1} y_{\beta-1})) \boldsymbol{x} \in \mathcal{B}_i, \boldsymbol{y} \in \mathbb{B}^{\beta} \}$						
$ \mathcal{C}_i = 2^{b_i} < 4^{\beta}$: number of codewords						

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

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

(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: $\forall u \in \mathcal{C}_i$

$$|w(u_{\mathrm{L}}) - w(u_{\mathrm{R}})| \leq \delta \ u_{L} = u_{0}^{\lfloor eta/2 \rfloor - 1}, u_{R} = u_{eta - \lfloor eta/2 \rfloor}^{eta - 1}$$

- (x) additional constraint (?)
- (4) generation fails if $|\mathcal{B}_i| < 2^{b_i'}$
- (5) full search: maximize $R_{
 m hd}$

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

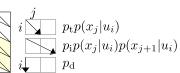
(*) motif: eliminated by mask (=QR code)

Metric

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







Local GC-balance

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