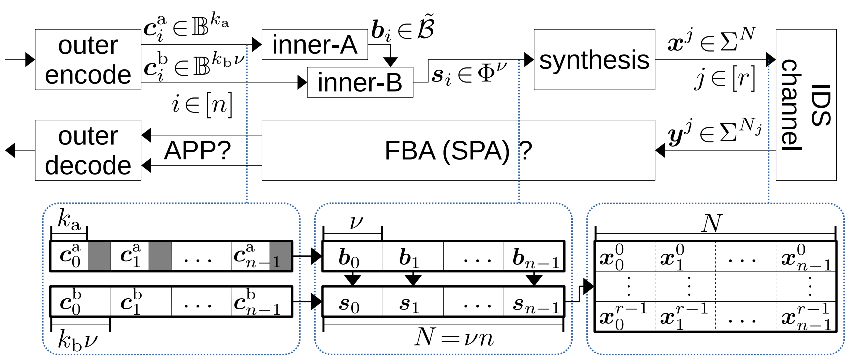


Alphabet: $\Sigma=\{0,1,2,3\}\leftrightarrow\{\mathrm{A,T,G,C}\}$
Code length: ν (even)
Block length: $N=\nu n$ symbol
(n codewords/block)

Constraints (binary):
run length: RL-ρ
local balance: LB-(ℓ,δ)
$$\left|w(b_i^{i+\ell-1})-\tfrac{\ell}{2}\right|\le\delta$$
indicator func: $\mathbb{1}_{\rho,\ell,\delta}(b)$

Binary constraint vector set

$\mathcal{B}\subset\mathbb{B}^\nu \quad (|\mathcal{B}|=2^{k_{\mathrm{a}}})$

- $\forall b\in\mathcal{B}, w(b)=\nu/2$ (LB: inside CW)
- $\forall b\in\mathcal{B}, \forall i\in[\nu]_{\mathrm{even}}, |\tilde{w}(b_i^{\nu-1})|\le 2\delta$ (LB: CW boundary)
- $\forall b\in\mathcal{B}, \forall i\in[\nu]_{\mathrm{odd}}, |\tilde{w}(b_i^{\nu-1})|\le 2\delta-1$ (LB: CW boundary)
- $\forall b\in\mathcal{B}, \lambda(b)\le\rho$ (RL: inside CW)
- $\forall b\in\mathcal{B}, \lambda(b,\nu-1)\le\rho-1$ (RL: CW boundary)
- $b\in\mathcal{B}\to\bar{b}\in\mathcal{B}$ (re-balance)
- $(01)^{\frac{\nu}{2}},(10)^{\frac{\nu}{2}}\notin\mathcal{B}$ (erasure symbol)

$\tilde{\mathcal{B}}=\mathcal{B}\cup\{(01)^{\frac{\nu}{2}},(10)^{\frac{\nu}{2}}\}$ (re-balance)

Composite symbol sets

$$\Phi_0\subseteq\left\{(\sigma_0,\sigma_1,\sigma_2,\sigma_3)\in\mathbb{Z}_{k+1}^4\,\middle|\,\sigma_0+\sigma_1=k,\sigma_2=\sigma_3=0\right\}\ \text{AT}$$
$$\Phi_1\subseteq\left\{(\sigma_0,\sigma_1,\sigma_2,\sigma_3)\in\mathbb{Z}_{k+1}^4\,\middle|\,\sigma_0=\sigma_1=0,\sigma_2+\sigma_3=k\right\}\ \text{GC}$$
$$|\Phi_0|=|\Phi_1|=2^{k_{\mathrm{b}}}\le k+1 \qquad \Phi=\Phi_0\cup\Phi_1$$

Inner codebook

$$\mathcal{C}\subseteq\left\{(c_0,\dots,c_{\nu-1})\,\middle|\,c_j\in\Phi_{b_j},(b_0,\dots,b_{\nu-1})\in\mathcal{B}\right\}$$

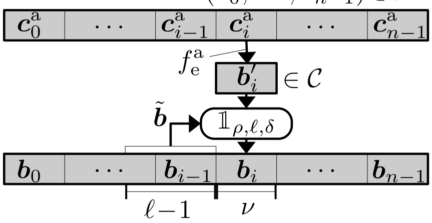
Mapping

$f_{\mathrm{e}}^{\mathrm{a}}:\mathbb{B}^{k_{\mathrm{a}}}\to\mathcal{B}$ (bijection)
$f_{\mathrm{e}}^{i}:\mathbb{B}^{k_{\mathrm{b}}}\to\Phi_i\ (i\in\mathbb{B})$ (bijection)
$\phi(x)=\mathbb{1}[x\in\{2,3\}] \quad (x\in\Sigma)$

Encoding

input: $c^{\mathrm{a}}=(c_0^{\mathrm{a}},\dots,c_{n-1}^{\mathrm{a}}) \quad c_i^{\mathrm{a}}\in\mathbb{B}^{k_{\mathrm{a}}}$
$c^{\mathrm{b}}=(c_0^{\mathrm{b}},\dots,c_{n-1}^{\mathrm{b}}) \quad c_i^{\mathrm{b}}\in\mathbb{B}^{k_{\mathrm{b}}\nu}$
output: $s=(s_0,\dots,s_{n-1}) \quad s_i\in\Phi^\nu$

(1) encode c^{a} to $b=(b_0,\dots,b_{n-1})\in\tilde{\mathcal{B}}^n$



$$b_i=\begin{cases} b_i' & (\mathbb{1}_{\rho,\ell,\delta}(\tilde{b},b_i')=1)\\ \bar{b}_i' & (\mathbb{1}_{\rho,\ell,\delta}(\tilde{b},\bar{b}_i')=1)\\ (b,\bar{b})^{\frac{\nu}{2}} & (\mathbb{1}_{\rho,\ell,\delta}(\tilde{b},(b,\bar{b})^{\frac{\nu}{2}})=1)\\ (\bar{b},b)^{\frac{\nu}{2}} & (\text{otherwise})\end{cases}$$

(2) encode (c^{b},b) to s

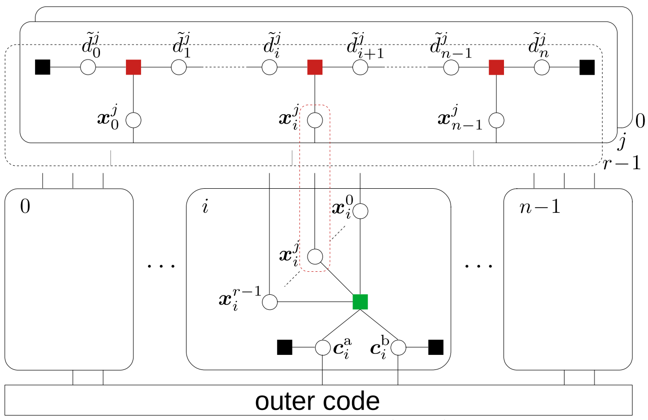
$$s_{i,t}=f_{\mathrm{e}}^{b_{i,t}}(c_{i,t}^{\mathrm{b}}) \qquad (i\in[n],t\in[\nu])$$
$$s_i=(s_{i,0},\dots,s_{i,\nu-1})\in\Phi^\nu$$
$$b_i=(b_{i,0},\dots,b_{i,\nu-1})\in\tilde{\mathcal{B}}$$
$$c_{i,t}^{\mathrm{b}}=[c_i^{\mathrm{b}}]_{tk_{\mathrm{b}}}^{tk_{\mathrm{b}}+k_{\mathrm{b}}-1}\in\mathbb{B}^{k_{\mathrm{b}}}$$

(ex) $\nu=6, |\mathcal{B}|=16, k=7, k_{\mathrm{a}}=4, k_{\mathrm{b}}=3$

```
B:0:001011  8:110100  Φ0:0:0700  Φ1: 0:0007
  1:001101  9:110010     1:1600      1:0016
  2:001110  A:110001     2:2500      2:0025
  3:010011  B:101100     3:3400      3:0034
  -:010101  -:101010     4:4300      4:0043
  4:010110  C:101001     5:5200      5:0052
  5:011001  D:100110     6:6100      6:0061
  6:011010  E:100101     7:7000      7:0070
  7:011100  F:100011   (Lee distance?)
```

```
c_i^a :0110(6)=011010
c_i^b :011(3)          s_i : (0,3)=3400
       100(4)                (1,4)=0043
       010(2)                (1,2)=0025
       011(3)                (0,3)=3400
       110(6)                (1,6)=0061
       100(4)                (0,4)=4300
```

Decoding



$j\in[r]$ $\qquad \tilde{d}_i^j=d_{i\nu}^j$ (symbol level)

$$p\left(y^0,\dots,y^{r-1},x^0,\dots,x^{r-1},\tilde{d}^0,\dots,\tilde{d}^{r-1},c^{\mathrm{a}},c^{\mathrm{b}}\right)$$
$$=p\left(y^0,\dots,y^{r-1}\middle|x^0,\dots,x^{r-1},\tilde{d}^0,\dots,\tilde{d}^{r-1},c^{\mathrm{a}},c^{\mathrm{b}}\right)\times$$
$$p\left(\tilde{d}^0,\dots,\tilde{d}^{r-1}\right)p\left(x^0,\dots,x^{r-1}|c^{\mathrm{a}},c^{\mathrm{b}}\right)p\left(c^{\mathrm{a}},c^{\mathrm{b}}\right)$$
$$=p(c^{\mathrm{a}})\,p(c^{\mathrm{b}})\prod_{j=0}^{r-1}p\left(y^j\middle|x^j,\tilde{d}^j\right)p\left(\tilde{d}^j\right)p\left(x^j|c^{\mathrm{a}},c^{\mathrm{b}}\right)$$
$$\simeq\left(\prod_{i=0}^{n-1}p(c_i^{\mathrm{a}})\,p(c_i^{\mathrm{b}})\right)\times$$
$$\prod_{j=0}^{r-1}\left(p(\tilde{d}_0^j)\prod_{i=0}^{n-1}\underline{p\left([y^j]_{i\nu+\tilde{d}_i^j}^{i\nu+\nu-1+\tilde{d}_{i+1}^j}\middle|x_i^j,\tilde{d}_i^j,\tilde{d}_{i+1}^j,\right)p\left(\tilde{d}_{i+1}^j\middle|\tilde{d}_i^j\right)}\ \underline{p\left(x_i^j\middle|c_i^{\mathrm{a}},c_i^{\mathrm{b}}\right)}\right)$$

$$\underline{\prod_{j=0}^{r-1}p\left(x_i^j\middle|c_i^{\mathrm{a}},c_i^{\mathrm{b}}\right)}=\prod_{j=0}^{r-1}p\left(b_i^j\middle|c_i^{\mathrm{a}}\right)\underbrace{p\left(x_i^j\middle|b_i^j,c_i^{\mathrm{b}}\right)}_{\text{(sim)}}=\prod_{j=0}^{r-1}p\left(b_i^j\middle|c_i^{\mathrm{a}}\right)\prod_{t=0}^{\nu-1}p\left(x_{i,t}^j\middle|b_{i,t}^j,c_{i,t}^{\mathrm{b}}\right)$$

$$b_i^j=(\phi(x_{i,0}^j),\dots,\phi(x_{i,\nu-1}^j))$$

$$p(x\,|\,b,c)=\sigma_x/k \quad (x\in\Sigma,b\in\mathbb{B},c\in\mathbb{B}^{k_{\mathrm{b}}})$$
$$(\sigma_0,\sigma_1,\sigma_2,\sigma_3)=f_{\mathrm{e}}^{b}(c)$$

non-uniform k

(example)

$\nu = 6, |\mathcal{B}| = 16, k = 7, k_{\mathrm{a}} = 4, k_{\mathrm{b}} = 3$

$\mathcal{B}$:		Φ_0:	Φ_1:
0:001011	8:110100	0:0700	0:0007
1:001101	9:110010	1:1600	1:0016
2:001110	A:110001	2:2500	2:0025
3:010011	B:101100	3:3400	3:0034
-:010101	-:101010	4:4300	4:0043
4:010110	C:101001	5:5200	5:0052
5:011001	D:100110	6:6100	6:0061
6:011010	E:100101	7:7000	7:0070
7:011100	F:100011		

(Lee distance?)

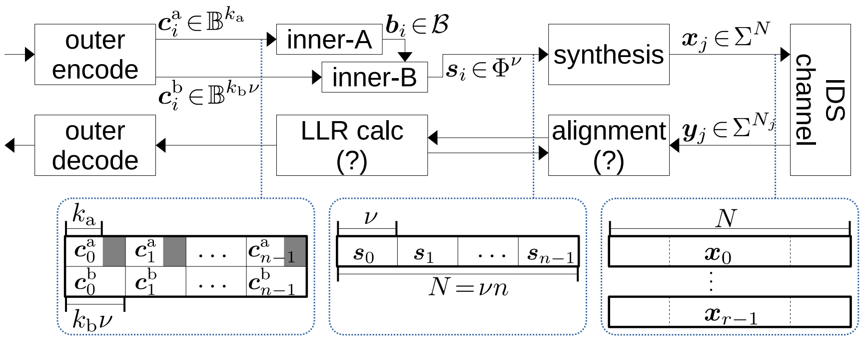
$\boldsymbol{c}_i^{\mathrm{a}}$: 0110(6)=011010

$\boldsymbol{c}_i^{\mathrm{b}}$: 011(3)
100(4)
010(2)
011(3)
110(6)
100(4)

$\boldsymbol{s}_i$: (0,3)=3400
(1,4)=0043
(1,2)=0025
(0,3)=3400
(1,6)=0061
(0,4)=4300

8 10

(A,T,G,C)
(0,1,2,3)



$i \in [n]$
$j \in [r]$
$N = \nu n$

Code length: ν (even)
Binary constraint vector set: $\mathcal{B} \subset \mathbb{B}^\nu \quad (|\mathcal{B}| \geq 2^{k_{\mathrm{a}}})$
Composite symbol set:

$\Phi_0 \subseteq \left\{(\sigma_0,\sigma_1,\sigma_2,\sigma_3) \in \mathbb{Z}_{k+1}^4 \,\middle|\, \sigma_0+\sigma_1=k, \sigma_2=\sigma_3=0\right\}$ AT

$\Phi_1 \subseteq \left\{(\sigma_0,\sigma_1,\sigma_2,\sigma_3) \in \mathbb{Z}_{k+1}^4 \,\middle|\, \sigma_0=\sigma_1=0, \sigma_2+\sigma_3=k\right\}$ GC

$|\Phi_0| = |\Phi_1| = 2^{k_{\mathrm{b}}} \leq k+1 \qquad \Phi = \Phi_0 \cup \Phi_1$

Inner codebook:

$\mathcal{C} \subseteq \left\{(c_0,\ldots,c_{\nu-1}) \,\middle|\, c_j \in \Phi_{b_j}, (b_0,\ldots,b_{\nu-1}) \in \mathcal{B}\right\}$

(example) $\nu=6, |\mathcal{B}|=16, k=7, k_{\mathrm{a}}=4, k_{\mathrm{b}}=3,$

$\mathcal{B}$:			Φ_0:	Φ_1:
	0:001011	8:110100	0:0700	0:0007
	1:001101	9:110010	1:1600	1:0016
	2:001110	A:110001	2:2500	2:0025
	3:010011	B:101100	3:3400	3:0034
	-:010101	-:101010	4:4300	4:0043
	4:010110	C:101001	5:5200	5:0052
	5:011001	D:100110	6:6100	6:0061
	6:011010	E:100101	7:7000	7:0070
	7:011100	F:100011		

Lee dist?

Encode:

$f_{\mathrm{e}}^{\mathrm{a}} : \mathbb{B}^{k_{\mathrm{a}}} \to \mathcal{B}$ (lossy)

$f_{\mathrm{e}}^{i} : \mathbb{B}^{k_{\mathrm{b}}} \to \Phi_i \ (i \in \mathbb{B})$ (bijection)

$f_{\mathrm{e}}(\boldsymbol{c}_i^{\mathrm{a}}, \boldsymbol{c}_i^{\mathrm{b}}) = \boldsymbol{s}_i = (s_{i,0},\ldots,s_{i,\nu-1}) : \mathbb{B}^{k_{\mathrm{a}}} \times \mathbb{B}^{k_{\mathrm{b}}\nu} \to \Phi^\nu$

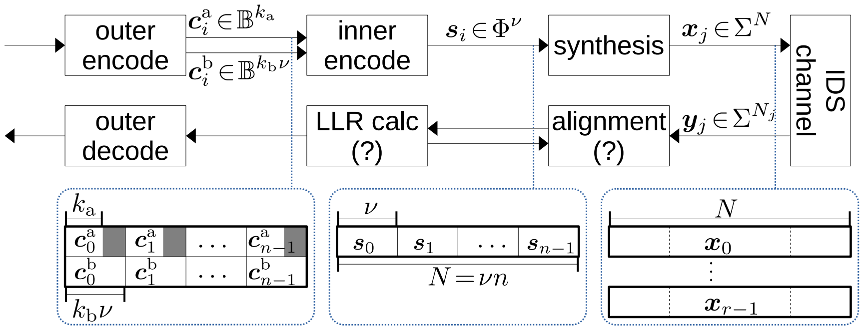
$s_{i,j} = f_{\mathrm{e}}^{b_i}(\boldsymbol{c}_{i,j}^{\mathrm{b}}) \qquad (b_0,\ldots,b_{\nu-1}) = f_{\mathrm{e}}^{\mathrm{a}}(\boldsymbol{c}_i^{\mathrm{a}})$

$\boldsymbol{c}_{i,j}^{\mathrm{b}} = [\boldsymbol{c}_i^{\mathrm{b}}]_{j\nu}^{j\nu+\nu-1}$

Decode:

8 10

(A,T,G,C)
(0,1,2,3)



$i \in [n]$
$j \in [r]$
$N = \nu n$

Code length: ν (even)
Binary constraint vector set: $\mathcal{B} \subset \mathbb{B}^\nu \quad (|\mathcal{B}| \geq 2^{k_{\mathrm{a}}})$
Composite symbol set:

$\Phi_0 \subseteq \left\{(\sigma_0, \sigma_1, \sigma_2, \sigma_3) \in \mathbb{Z}_{k+1}^4 \,\middle|\, \sigma_0 + \sigma_1 = k, \sigma_2 = \sigma_3 = 0\right\}$ AT

$\Phi_1 \subseteq \left\{(\sigma_0, \sigma_1, \sigma_2, \sigma_3) \in \mathbb{Z}_{k+1}^4 \,\middle|\, \sigma_0 = \sigma_1 = 0, \sigma_2 + \sigma_3 = k\right\}$ GC

$|\Phi_0| = |\Phi_1| = 2^{k_{\mathrm{b}}} \leq k+1 \qquad \Phi = \Phi_0 \cup \Phi_1$

Inner codebook:

$\mathcal{C} \subseteq \left\{(c_0, \ldots, c_{\nu-1}) \,\middle|\, c_j \in \Phi_{b_j}, (b_0, \ldots, b_{\nu-1}) \in \mathcal{B}\right\}$

(example) $\nu = 6, |\mathcal{B}| = 16, k = 7, k_{\mathrm{a}} = 4, k_{\mathrm{b}} = 3,$

Lee dist?

$\mathcal{B}$:		Φ_0 :	Φ_1 :
001011	110100	0700	0007
001101	110010	1600	0016
001110	110001	2500	0025
010011	101100	3400	0034
010101	101010	4300	0043
010110	101001	5200	0052
011001	100110	6100	0061
011010	100101	7000	0070
011100	100011		

Encode:

$f_{\mathrm{e}}^{\mathrm{a}} : \mathbb{B}^{k_{\mathrm{a}}} \to \mathcal{B}$
$f_{\mathrm{e}}^{i} : \mathbb{B}^{k_{\mathrm{b}}} \to \Phi_i \ (i \in \mathbb{B})$ bijection

$f_{\mathrm{e}}(\boldsymbol{c}_i^{\mathrm{a}}, \boldsymbol{c}_i^{\mathrm{b}}) = \boldsymbol{s}_i = (s_{i,0}, \ldots, s_{i,\nu-1}) : \mathbb{B}^{k_{\mathrm{a}}} \times \mathbb{B}^{k_{\mathrm{b}}\nu} \to \Phi^\nu$

$s_{i,j} = f_{\mathrm{e}}^{b_i}(\boldsymbol{c}_{i,j}^{\mathrm{b}}) \qquad (b_0, \ldots, b_{\nu-1}) = f_{\mathrm{e}}^{\mathrm{a}}(\boldsymbol{c}_i^{\mathrm{a}})$

$\boldsymbol{c}_{i,j}^{\mathrm{b}} = [\boldsymbol{c}_i^{\mathrm{b}}]_{j\nu}^{j\nu+\nu-1}$

flip (RL, LB)

Decode:

```
         001011 001101 001110 ...110100
base     AAGAGG AAGGAG AAGGGA    GGAGAA
         AAGAGC AAGGAC AAGGGT    GGAGAT
         AAGACG AAGGTG AAGGCA    GGAGTA   64
         ...    ...    ...       ...
         TTCTCC TTCCTC TTCCCT    CCTCTT
composite
                                          ??
```

multi-base composite (Manchester coding)

$\nu = 1:$ (A,C,G,T)
$\nu = 2:$ (AA,AC,AG,AT, CA,CC,CG,CT, GA,GC,GG,GT, TA,TC,TG,TT)
$\nu = 3:$ (AAA, AAC, ..., TTT)

```
         001011 001101 001110 ...110100
base     AAGAGG AAGGAG AAGGGA    GGAGAA
         AAGAGC AAGGAC AAGGGT    GGAGAC
         AAGACG AAGGTG AAGGCA    GGAGCA   64
         ...    ...    ...       ...
         CCTCTT CCTTCT CCTTTC
composite
```