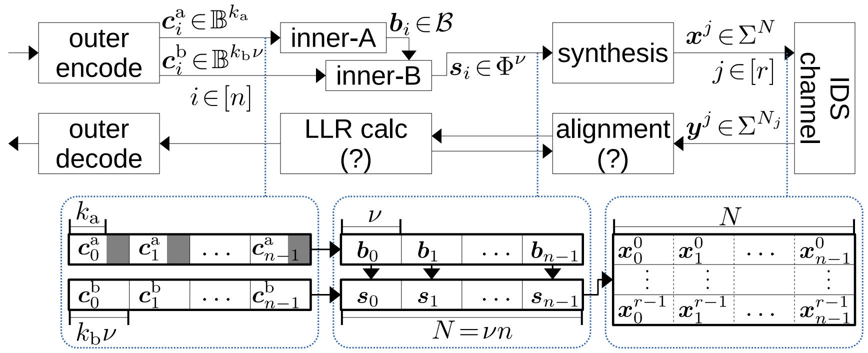
8 10

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



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

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

```
B:                 Φ0:     Φ1:
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?)

[Binary constraint vector set]

$\mathcal{B} \subset \mathbb{B}^{\nu} \quad (|\mathcal{B}| \geq 2^{k_{\mathrm{a}}}, \mathcal{B} \cap \{(01)^{\frac{\nu}{2}}, (10)^{\frac{\nu}{2}}\} = \phi)$

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

[Composite symbol sets]

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

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

[Encode]

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

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

input: $\boldsymbol{c}^{\mathrm{a}} = (\boldsymbol{c}_0^{\mathrm{a}},\ldots,\boldsymbol{c}_{n-1}^{\mathrm{a}}) \quad \boldsymbol{c}_i^{\mathrm{a}} \in \mathbb{B}^{k_{\mathrm{a}}}$

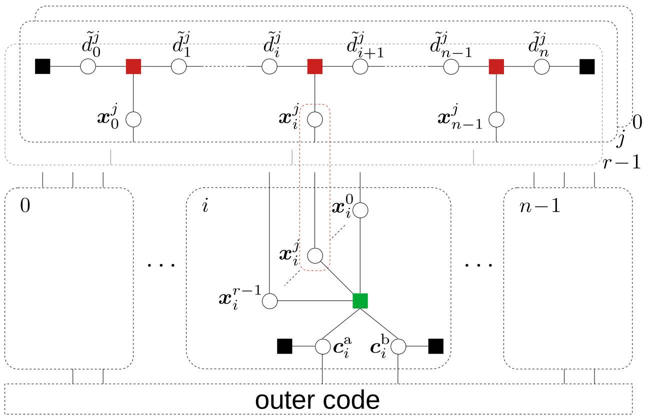
$\boldsymbol{c}^{\mathrm{b}} = (\boldsymbol{c}_0^{\mathrm{b}},\ldots,\boldsymbol{c}_{n-1}^{\mathrm{b}}) \quad \boldsymbol{c}_i^{\mathrm{b}} \in \mathbb{B}^{k_{\mathrm{b}}\nu}, \boldsymbol{c}_{i,j}^{\mathrm{b}} = [\boldsymbol{c}_i^{\mathrm{b}}]_{j\nu}^{j\nu+\nu-1}$

output: $\boldsymbol{s} = (\boldsymbol{s}_0,\ldots,\boldsymbol{s}_{n-1}) \quad \boldsymbol{s}_i = (s_{i,0},\ldots,s_{i,\nu-1}) \in \Phi^{\nu}$

(1) $\boldsymbol{b} = (\boldsymbol{b}_0,\ldots,\boldsymbol{b}_{n-1}) = f_{\mathrm{e}}^{\mathrm{a}}(\boldsymbol{c}^{\mathrm{a}}) \quad \boldsymbol{b}_i = (b_{i,0},\ldots,b_{i,\nu-1}) \in \tilde{\mathcal{B}}$

(2) $s_{i,j} = f_{\mathrm{e}}^{b_i}(\boldsymbol{c}_{i,j}^{\mathrm{b}}) \quad (i \in [n], j \in [\nu])$

[Decode]



$\tilde{d}_i^j = d_{i\nu}^j$ (symbol level)

$j \in [r]$

$$p\left(\boldsymbol{y}^0,\ldots,\boldsymbol{y}^{r-1},\boldsymbol{x}^0,\ldots,\boldsymbol{x}^{r-1},\tilde{\boldsymbol{d}}^0,\ldots,\tilde{\boldsymbol{d}}^{r-1},\boldsymbol{c}^{\mathrm{a}},\boldsymbol{c}^{\mathrm{b}}\right)$$

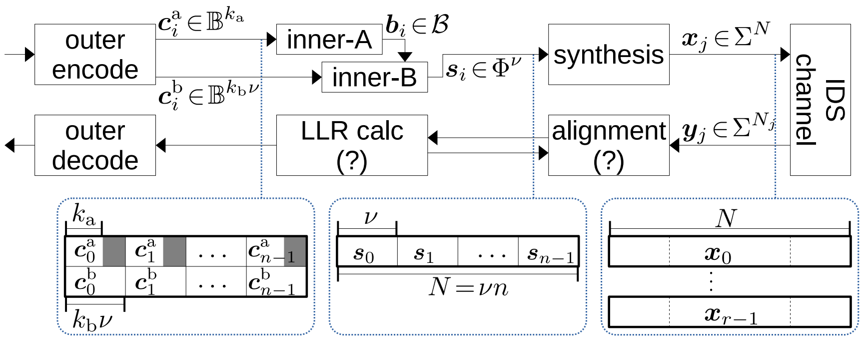
$$= p\left(\boldsymbol{y}^0,\ldots,\boldsymbol{y}^{r-1}\middle|\boldsymbol{x}^0,\ldots,\boldsymbol{x}^{r-1},\tilde{\boldsymbol{d}}^0,\ldots,\tilde{\boldsymbol{d}}^{r-1},\boldsymbol{c}^{\mathrm{a}},\boldsymbol{c}^{\mathrm{b}}\right) p\left(\tilde{\boldsymbol{d}}^0,\ldots,\tilde{\boldsymbol{d}}^{r-1}\right) p\left(\boldsymbol{x}^0,\ldots,\boldsymbol{x}^{r-1}\middle|\boldsymbol{c}^{\mathrm{a}},\boldsymbol{c}^{\mathrm{b}}\right) p\left(\boldsymbol{c}^{\mathrm{a}},\boldsymbol{c}^{\mathrm{b}}\right)$$

$$= p\left(\boldsymbol{c}^{\mathrm{a}}\right) p\left(\boldsymbol{c}^{\mathrm{b}}\right) \prod_{j=0}^{r-1} p\left(\boldsymbol{y}^j\middle|\boldsymbol{x}^j,\tilde{\boldsymbol{d}}^j\right) p\left(\tilde{\boldsymbol{d}}^j\right) p\left(\boldsymbol{x}^j\middle|\boldsymbol{c}^{\mathrm{a}},\boldsymbol{c}^{\mathrm{b}}\right)$$

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

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: $\mathcal{B} \subset \mathbb{B}^{\nu}$ $(|\mathcal{B}| \geq 2^{k_{\mathrm{a}}})$
Composite symbol set:

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

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

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

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)

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

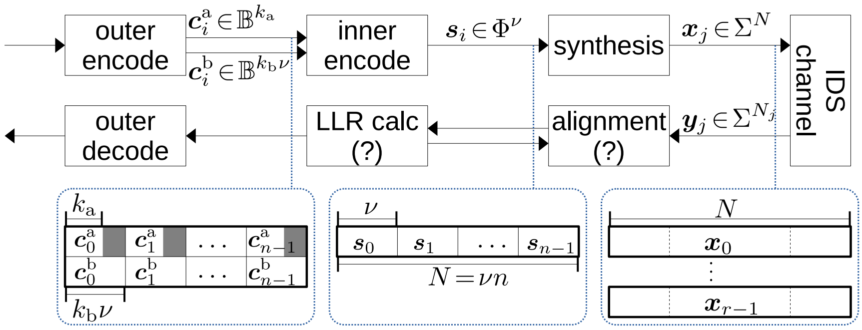
$$s_{i,j} = f_{\mathrm{e}}^{b_i}(c_{i,j}^{\mathrm{b}}) \qquad (b_0,\ldots,b_{\nu-1}) = f_{\mathrm{e}}^{\mathrm{a}}(c_i^{\mathrm{a}})$$

$$c_{i,j}^{\mathrm{b}} = [c_i^{\mathrm{b}}]_{j\nu}^{j\nu+\nu-1}$$

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:

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

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

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

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:

$$\left.\begin{array}{l} f_{\mathrm{e}}^{\mathrm{a}} : \mathbb{B}^{k_{\mathrm{a}}} \to \mathcal{B} \\ f_{\mathrm{e}}^{i} : \mathbb{B}^{k_{\mathrm{b}}} \to \Phi_i \ (i \in \mathbb{B}) \end{array}\right] \text{bijection}$$

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

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

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

flip (RL, LB)

Decode:

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