Chapter

The COVID-19 DNA-RNA Genetic Code Analysis Using Information Theory of Double Stochastic Matrix

Sung Kook Lee and Moon Ho Lee

Abstract

We present a COVID-19 DNA-RNA genetic code where $A = T = U = 31\%$ and $C = G = 19\%$, which has been developed from a base matrix $[C\,U; A\,G]$ where C, U, A, and G are RNA bases while C, U, A, and T are DNA bases that E. Chargaff found them complementary like $A = T = U = 30\%$, and $C = G = 20\%$ from his experimental results, which implied the structure of DNA double helix and its complementary combination. Unfortunately, they have not been solved mathematically yet. Therefore, in this paper, we present a simple solution by the information theory of a doubly stochastic matrix over the Shannon symmetric channel as well as prove it mathematically. Furthermore, we show that DNA-RNA genetic code is one kind of block circulant Jacket matrix. Moreover, general patterns by block circulant, upper-lower, and left-right scheme are presented, which are applied to the correct communication as well as means the healthy condition because it perfectly consists of 4 bases. Henceforth, we also provide abnormal patterns by block circulant, upper-lower, and left-right scheme, which cover the distorted signal as well as COVID-19.

Keywords: COVID-19 DNA-RNA, E. Chargaff, DNA-RNA genetic code, double stochastic matrix, symmetric channel, block circulant jacket matrix, general pattern, abnormal pattern

1. Introduction

In 1950, Chargaff's two rules [1] were presented. One is that the percentage of adenine is identical to that of thymine as well as the percentage of guanine is identical to that of cytosine, which gives a hint of the composition of the base pair for the double-strand DNA molecule. The other is that base complementarity is effective for each DNA strand, which gives an explanation for the overall characteristics of fundamental bases. To make an example of COVID-19 DNA, its four bases are satisfied with these two rules analogous to $A = T = 31\%$ and $C = G = 19\%$. In 1953, it was discovered that DNA has a double helix structure [2, 3], which results in an optimal and economical genetic code [4].

A RNA base matrix $[C\,U; A\,G]$ was based on stochastic matrices [5], which results in the genetic code [6, 7]. A symmetric capacity is calculated by applying the Markov process to these doubly stochastic matrices, which suggested the symmetry between

Shannon [8] and RNA stochastic transition matrix $[C\,U; A\,G]$, which is defined as below. A square matrix of $\mathbf{P} = \left(p_{ij}\right)$ is stochastic, whose entries are positive as well as its sum in rows and columns is equal to one or constant. In other words, if the sum of all its elements in rows and columns is equal to one or invariable, it is double stochastic, which is able to describe the time-invariant binary symmetric channel. For the input x_n and the output x_{n+1}, two states e_0 and e_1 are able to depict Markov processes on an individual basis, which are indicated by two binary symbols "0" and "1", accordingly. The output signal is affected by the input signal whose information is fed into given a certain error probability. Assume that these channel probabilities α and β are less than a half, whose error probabilities have been kept steady over a time-variant channel for a wide variety of transmitted symbols such as

$$P\{\mathrm{x}_{n+1} = 1|\mathrm{x}_n = 0\} = p_{01} = \alpha, P\{\mathrm{x}_{n+1} = 0|\mathrm{x}_n = 1\} = p_{10} = \beta. \tag{1}$$

In addition, its Markov chain is homogeneous. $\mathbf{P}$ represents a 2×2 homogeneous probability transition matrix defined as

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}_{p=0.5} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{2}$$

whose two error probabilities are identical similarly to $\alpha = \beta = p$ over a binary symmetric channel. This paper proceeds as below. First of all, we derive the RNA stochastic entropy by applying it to the Shannon entropy in Section 2. Next, we make an estimate of the variance of RNA in Section 3. Then, the binary symmetric channel entropy is derived in Section 4. Henceforth, two user capacity is made an estimate of over symmetric interference channel in Section 5. Afterward, the construction scheme is proposed, which is enabled to create RNA genetic codes in Section 6. Later, a symmetric genetic Jacket block matrix is examined in Section 7. Hereupon, general

Organism	Taxon	%A	%G	%C	%T	A / T	G / C	%GC	%AT
Maize	*Zea*	26.8	22.8	23.2	27.2	0.99	0.98	46.1	54.0
Octopus	*Octopus*	33.2	17.6	17.6	31.6	1.05	1.00	35.2	64.8
Chicken	*Gallus*	28.0	22.0	21.6	28.4	0.99	1.02	43.7	56.4
Rat	*Rattus*	28.6	21.4	20.5	28.4	1.01	1.00	42.9	57.0
Human	*Homo*	29.3	20.7	20.0	30.0	0.98	1.04	40.7	59.3
Grasshopper	*Orthoptera*	29.3	20.5	20.7	29.3	1.00	0.99	41.2	58.6
Sea urchin	*Echinoidea*	32.8	17.7	17.3	32.1	1.02	1.02	35.0	64.9
Wheat	*Triticum*	27.3	22.7	22.8	27.1	1.01	1.00	45.5	54.4
Yeast	*Saccharomyces*	31.3	18.7	17.1	32.9	0.95	1.09	35.8	64.4
E. coli	*Escherichia*	24.7	26.0	25.7	23.6	1.05	1.01	51.7	48.3
φX174	*PhiX174*	24.0	23.3	21.5	31.2	0.77	1.08	44.8	55.2
Covid-19	*SARS-CoV-2*	29.9	19.6	18.4	32.1	0.93	1.07	38.0	62.0

Table 1.
Ratio of bases [1, 9–11].

patterns of block circulant symmetric genetic Jacket matrices are looked into in Section 8. In the end, this paper comes to a conclusion in Section 9.

Table 1 makes the description of the ratio of bases for several organisms [1, 9–11], which shows that the ratios are constant among the species.

2. Analytical approach to RNA stochastic entropy

In [1, 5, 12, 13], stochastic complementary RNA bases are given for the genetic code. On the assumption that $C = G = 19\%$, $A = T = U = 31\%$, $\mathbf{P}$ denotes the transition channel matrix expressed by

$$\mathbf{P} = \begin{bmatrix} C & U \\ A & G \end{bmatrix} = \begin{bmatrix} 0.19 & 0.31 \\ 0.31 & 0.19 \end{bmatrix}. \tag{3}$$

On the condition that the RNA base matrix $[C\,U; A\,G]$ for the Markov process described by two independent probabilities of its corresponding source varies from $0.19p$ to $0.31p$, the transition channel matrix $\mathbf{P}$ is defined by

$$\mathbf{P} = \begin{bmatrix} 0.19p & 1-0.19p \\ 1-0.19p & 0.19p \end{bmatrix} = \begin{bmatrix} 0.5 & 1-0.5 \\ 1-0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}. \tag{4}$$

By comparison with Eq. (12), we have.

$$0.19p = 1 - 0.19p \tag{5}$$

where p is 2.631.
Applying in a similar fashion to the rest of (4),

$$\mathbf{P} = \begin{bmatrix} 0.31p & 1-0.31p \\ 1-0.31p & 0.31p \end{bmatrix} = \begin{bmatrix} 0.500 & 1-0.500 \\ 1-0.500 & 0.500 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \tag{6}$$

where $0.31p$ = 1-$0.31p$, where p is 1.613.
In order to make a double stochastic matrix by adding (6) to (4),

$$2\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{7}$$

Applying in a similar way to (3),

$$2\mathbf{P} = 2\begin{bmatrix} C & U \\ A & G \end{bmatrix} = 2\begin{bmatrix} 0.19 & 0.31 \\ 0.31 & 0.19 \end{bmatrix} = \begin{bmatrix} 0.38 & 0.62 \\ 0.62 & 0.38 \end{bmatrix}. \tag{8}$$

If P is a random variable for source probability p corresponding to the first symbol event, we reach the entropy function [8] represented by

$$H_2(P) = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right). \tag{9}$$

P	$-\log_2 p$	$-p\log_2 p$	$H_2(p)$
0.3800	1.3959	0.5305	0.9580
0.3900	1.3585	0.5298	0.9648
0.4000	1.3219	0.5288	0.9710
0.4100	1.2863	0.5274	0.9765
0.4200	1.2515	0.5256	0.9815
0.4300	1.2176	0.5236	0.9858
0.4400	1.1844	0.5211	0.9896
0.4500	1.1520	0.5184	0.9928
0.4600	1.1203	0.5153	0.9954
0.4700	1.0893	0.5120	0.9974
0.4800	1.0589	0.5083	0.9988
0.4900	1.0291	0.5043	0.9997
0.5000	1.0000	0.5000	1.0000
0.5100	0.9714	0.4954	0.9997
0.5200	0.9434	0.4906	0.9988
0.5300	0.9159	0.4854	0.9974
0.5400	0.8890	0.4800	0.9954
0.5500	0.8625	0.4744	0.9928
0.5600	0.8365	0.4684	0.9896
0.5700	0.8110	0.4623	0.9858
0.5800	0.7859	0.4558	0.9815
0.5900	0.7612	0.4491	0.9765
0.6000	0.7370	0.4422	0.9710
0.6100	0.7131	0.4350	0.9648
0.6200	0.6897	0.4276	0.9580

Table 2.
Shannon entropy for probability p.

The last column of **Table 2** shows the result of Eq. (9). **Figure 1** portrays the curve of Shannon and RNA Entropy. Make a mental note to make sure that a vertical tangent can be drawn when $p = 0$ and $p = 1$ on account of the fact that

$$\frac{d}{dp}\left[p\log_2\left(\frac{1}{p}\right)+(1-p)\log_2\left(\frac{1}{1-p}\right)\right]=\left[\log_2\left(\frac{1}{p}\right)-1-\log_2\left(\frac{1}{1-p}\right)+1\right]\log_2 e$$
$$=\log_2\left(\frac{1}{p}\right)-\log_2\left(\frac{1}{1-p}\right)=0, \tag{10}$$

which is maximized when p reaches a half because its derivative becomes 0.

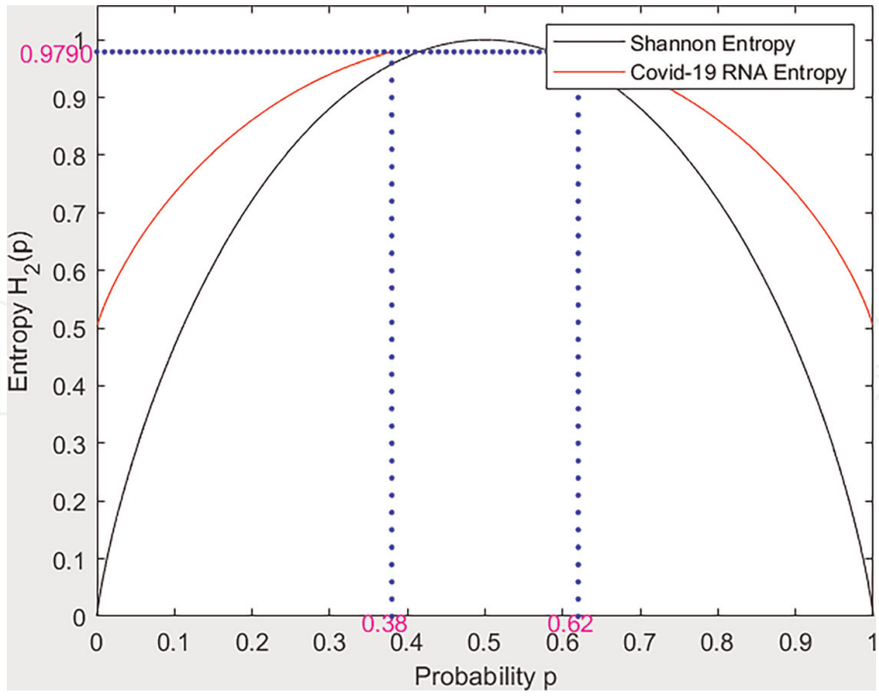

Figure 1.
Comparison between Shannon and RNA entropy for probability p.

Therefore,

$$\log_2\left(\frac{1}{p}\right) - \log_2\left(\frac{1}{1-p}\right) = 0 \Rightarrow \left(\frac{1}{p}\right) - \left(\frac{1}{1-p}\right) = 0. \tag{11}$$

Then, we reach

$$p = 1 - p \Rightarrow p = \frac{1}{2}. \tag{12}$$

For the RNA base matrix $[C\ U; A\ G]$, its symmetric entropy is calculated as

$$H_2(P)_{RNA} = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right) = 0.9790, \tag{13}$$

when p is either 0.38 or 0.62. By the way, the Shannon entropy is calculated as

$$H_2(P)_{Shannon} = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right) = 1, \tag{14}$$

when p reaches a half.

Table 2 shows Shannon Entropy for probability p over a binary symmetric channel.

Figure 1 gives a comparison between Shannon and RNA Entropy for probability p under the RNA base matrix $[C\ U; A\ G]$.

3. Derivation of variance for the RNA base matrix $[C\,U; A\,G]$

The variance for RNA random variable X is denoted by $V(X)$ is the square of the mean, which is expressed by

$$E\{X\} = a = 0.5. \tag{15}$$

Therefore, for a random variable X, the variance is obtained such as

$$\begin{aligned} V(X) &= E\left\{(X-a)^2\right\} = E\{X^2\} - 2aE\{X\} + E\{a^2\} \\ &= E\{X^2\} - 2a^2 + a^2 = E\{X^2\} - a^2 = \sigma^2. \end{aligned} \tag{16}$$

Case I. Upper source probability 0.62

$$\sigma^2_{upper} = (0.62)^2 - (0.5)^2 = 0.13. \tag{17}$$

Case II. Lower source probability 0.38

$$\sigma^2_{lower} = (0.5)^2 - (0.38)^2 = 0.10. \tag{18}$$

If X_1 and X_2 are the independent random variables, on an individual basis, its expectation and variance are

$$E\{X_1\} = a_1, \quad V\{X_1\} = \sigma_1^2. \tag{19}$$

$$E\{X_2\} = a_2, \quad V\{X_2\} = \sigma_2^2. \tag{20}$$

Therefore, we reach

$$E\{(X_1 - a_1)(X_2 - a_2)\} = E\{(X_1 - a_1)\}E\{(X_2 - a_2)\} = 0. \tag{21}$$

Assuming that X_1 and X_2 are independent random variables, the sum of its variances is calculated as

$$\begin{aligned} V\{X_1 + X_2\} &= E\left\{(X_1 + X_2 - a_1 - a_2)^2\right\} \\ &= E\left\{(X_1 - a_1)^2\right\} + 2E\{(X_1 - a_1)(X_2 - a_2)\} + E\left\{(X_2 - a_2)^2\right\} \\ &= V\{X_1\} + V\{X_2\} = \sigma_1^2 + \sigma_2^2 = 0.13 + 0.10 = 0.23, \end{aligned} \tag{22}$$

which is approximately 23% corresponding to the difference between A = U and C = G. It means that RNA entropy cannot reach the Shannon entropy because the probabilities of its bases are 23% away from a half that is exactly identical to the sum of its variances.

4. RNA complement base matrix $[C\,U; A\,G]$ for symmetric noise immune-free channel

If over a noise immune-free binary symmetric channel the bases of RNA genetic code $[C\,U; A\,G]$ are complementary such as $C = U$ and $A = G$, the conditional

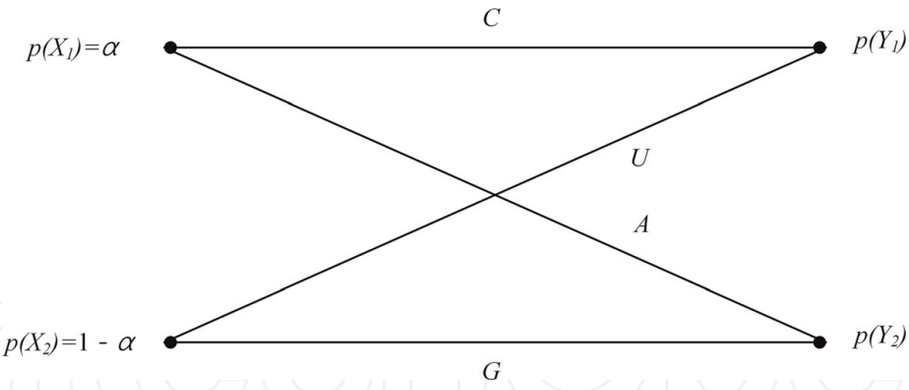

Figure 2.
Complementary bases of RNA genetic code [C U; A G] over noise immune-free binary symmetric channel.

probability $P(b_j|a_i) = P_{i,j}$ makes description of this channel, whose maximum amount of information can be transmitted as depicted in **Figure 2**. On the assumption that C and G are one's complement of its corresponding error probability as well as A and U are interference signals, the matrix [8] for this channel is made description of by

$$[p(X)]_{1\times 2}[P]_{2\times 2} = [\alpha \;\; 1-\alpha]\begin{bmatrix} C & U \\ A & G \end{bmatrix} = [p(Y)]_{1\times 2} = [\,p(Y_1) \quad p(Y_2)\,]. \tag{23}$$

Under the condition that p and 1-p are the selection probability $(\alpha = 0)$ and $(\alpha = 1)$ over the uniform channel on an individual basis, the mutual information is defined by

$$I(X;Y) = H(Y) - H(Y|X). \tag{24}$$

From Eq. (23), we are confronted with

$$[\,\alpha \quad 1-\alpha\,]\begin{bmatrix} -C\log_2 C & -U\log_2 U \\ -A\log_2 A & -G\log_2 G \end{bmatrix} = [\,\alpha \quad 1-\alpha\,]\begin{bmatrix} -U\log_2 U & -C\log_2 C \\ G\log_2 G & -A\log_2 A \end{bmatrix}, \tag{25}$$

where

$$\begin{aligned} H(Y|X) &= -\alpha C\log_2 C - \alpha A\log_2 A - (1-\alpha)U\log_2 U - (1-\alpha)G\log_2 G \\ &= -U\log_2 U - G\log_2 G = -C\log_2 C - A\log_2 A = 0.9790, \end{aligned} \tag{26}$$

where $A = U = 0.31$ and $C = G = 0.19$.
Therefore, its capacity is derived as

$$C_{RNA} = \max I(X;Y)\big|_{p=0.38 \text{ or } 0.62} = H(Y) - H(Y|X) = 1 - 0.9790 = 0.021, \tag{27}$$

i.e. $H(Y) = -p\log_2 p - (1-p)\log_2(1-p) = -0.38\log_2 0.38 - 0.62\log_2 0.62 = 1.$
while Shannon capacity is derived as

$$C_{Shannon} = \max I(X;Y)\big|_{p=0.5} = H(Y) - H(Y|X) = 1 - 1 = 0. \tag{28}$$

In **Figure 3**, we compare Shannon and RNA capacity for probability p. As forementioned in Section 3, if only if under the ideal circumstance, Shannon capacity can

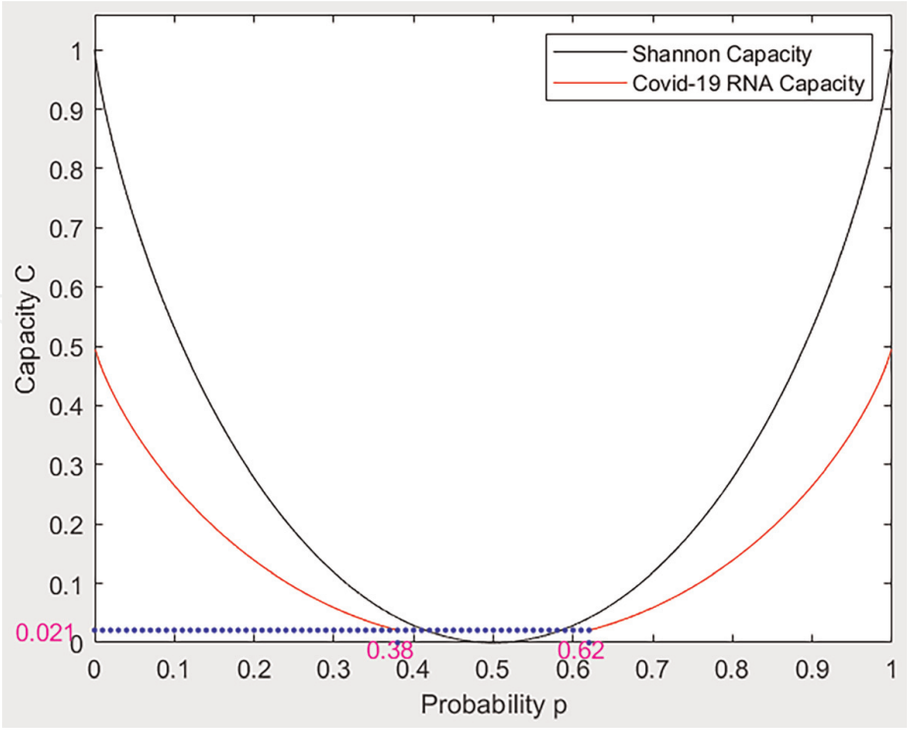

Figure 3.
Shannon and RNA capacity vary with probability p.

be reached. In other words, the difference between Shannon and RNA capacity exists, which is identical to the sum of variances of RNA base random variables because they are unable to become a half over a symmetric channel.

5. Two user capacity over symmetric interference channel

Figure 4 makes the description of the environment of the binary symmetric channel with the RNA base matrix $[C\ U; A\ G]$ as well as that of the symmetric interference channel for two users where two independent messages W_1 and W_2 with the common message set W_i are transmitted. Assume that $C = G = 19\%$ and $A = U = 31\%$ where $C = H^{11}$ is the direct signal and its corresponding interference signal is $U = H^{12}$ for Y_1. Analogously, the direct signal for the second user Y_2 is $G = H^{22}$ and its corresponding interference signal is $A = H^{21}$.

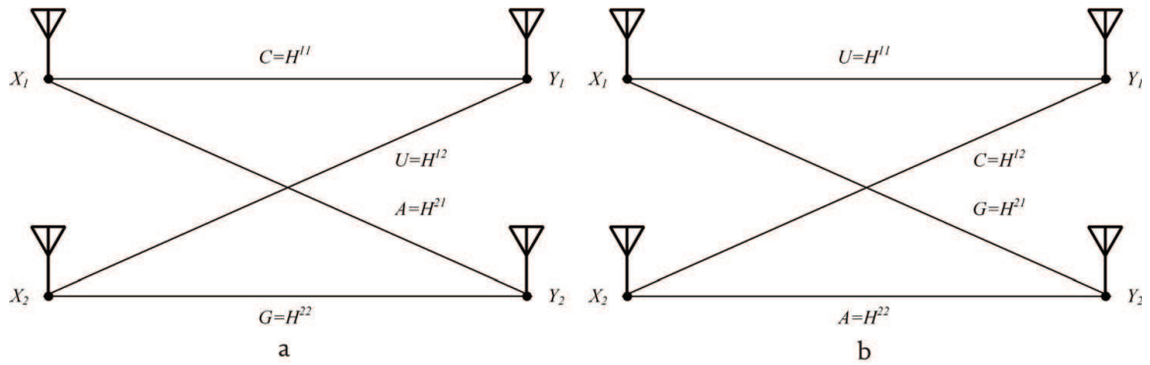

Figure 4.
Two-user symmetric Interference Channel. (a) Strong Interference Channel. (b) Weak Interference Channel.

$$H^{11} = H^{12} = h_d\sqrt{P_{SNR}},$$
$$H^{12} = H^{21} = h_c\sqrt{P_{SNR}}.$$

The relationship between the input and output for two user symmetric channel is described as follows [14],

$$Y_1 = h_d\sqrt{P_{SNR}}X_1 + h_c\sqrt{P_{SNR}^{\alpha}}X_2 + Z_1, \tag{29}$$
$$Y_2 = h_c\sqrt{P_{SNR}^{\alpha}}X_1 + h_d\sqrt{P_{SNR}}X_2 + Z_2, \tag{30}$$

where the powers of input symbols X_1, X_2, and additive white Gaussian noise (AWGN) terms Z_1 and Z_2 are normalized to unity. Analogous to the definition of the degree of freedom (*DoF*), the total *GDoF* metric $d(\alpha)$ is defined as

$$d(\alpha) = \lim_{P_{SNR}\to\infty}\frac{C(P_{SNR},\alpha)}{\log(P_{SNR})}, \tag{31}$$

where C (P_{SNR}, α) is the sum-capacity parameterized by P_{SNR} and α. Here α is the ratio (on the decibel scale) of cross channel strength compared to straight channel strength and P_{SNR} indicates the ratio (on the decibel scale) of signal to the noise. Importantly, in order to find the achievable *DoF*, take the limit of Eq. (31) by letting P_{SNR} go to infinity. Make a mental note of the *DoF* metric resembling to that at the point α =1. Thus, the *GDoF* curve gives a significant hint for optimal interference management strategies, which has been made use of most successfully to estimate the capacity of two-user interference channel to contain a constant gap in [14]. To take an example, for RNA genetic code, assuming that its bases $C = G$ = 19% and $A = T = U$ = 31%, this symmetric interference channel for two users can be analyzed in strong and weak interference region as below. The noise immune channel is described as below where X_1 and X_2 denote the input symbols while Y_1 and Y_2 denote the output symbols

$$Y_1 = CX_1 + UX_2, \tag{32}$$
$$Y_2 = GX_1 + AX_2. \tag{33}$$

Case 1. Strong Interference region.

Figure 4 (a) makes the description of the channel in a strong interference regime, where its receivers have to try to decode the interfering signal in order to recover its desired signal. The general condition for a strong interference signal is represented by,

$$C < A, U > G. \tag{34}$$

Regretfully, it is still challenging to propose the scheme achieving a symmetric rate as well as being upper-bounded unlike in the weak interference region.

Case 2. Weak Interference region.

Figure 4 (b) makes the description of the channel in a very weak interference regime, where its receivers do not need to try to decode any portion of the interference signal by regarding it as noise. This scheme is enabled to achieve a symmetric rate per user as below [7],

$$R = \min\left\{\frac{1}{2}\log(1 + INR + SNR) + \frac{1}{2}\log\left(2 + \frac{SNR}{INR}\right) - 1,\ \log\left(1 + INR + \frac{SNR}{INR}\right) - 1\right\}. \tag{35}$$

The upper bound on the symmetric capacity is,

$$C_{Sym} \le \min\left(\frac{1}{2}\log\left(1+SNR\right)+\frac{1}{2}\log\left(1+\frac{SNR}{1+INR}\right),\ \log\left(1+INR+\frac{SNR}{1+INR}\right)\right). \tag{36}$$

Letting $A = T = U = 31\%$, $C = G = 19\%$, i.e. $INR = 31$ and $SNR = 19$, we are confronted with the symmetric achievable rate such as

$$\begin{aligned} R &= \min\left\{\frac{1}{2}\log_2(1+31+19)+\frac{1}{2}\log_2\left(2+\frac{19}{31}\right)-1,\ \log_2\left(1+31+\frac{19}{31}\right)-1\right\} \\ &= \min\left\{2.83+0.69-1, 5.02-1\right\} = \min\left\{2.53, 4.02\right\} = 2.52. \end{aligned} \tag{37}$$

Analogously, the symmetric capacity is made the description of by

$$\begin{aligned} C_{sym} &\le \min\left\{\frac{1}{2}\log_2(1+19)+\frac{1}{2}\log_2\left(1+\frac{19}{31}\right),\ \log_2\left(1+31+\frac{19}{31}\right)\right\} \\ &\le \min\left\{2.16+0.34, 5.02\right\} \le \min\left\{2.50, 5.02\right\} = 2.50. \end{aligned} \tag{38}$$

Following the above steps, in a weak interference regime, by treating interference as noise, the symmetric capacity is close to its achievable capacity such as

$$C_{sym} = R. \tag{39}$$

Figure 5 makes the description of the weak and strong interference region where the leftmost indicates a very weak interference region while the rightmost suggests a very strong interference region.

Analysis:

In 1948, Shannon proposed the code generation method by exploiting the random codebook in point-to-point communication with inverse Gaussian distribution (Gaussian distribution variance towards infinity is called inverse Gaussian) to achieve the channel capacity, which is described as follows [8],

$$C = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right), \tag{40}$$

where the signal power is S and the noise power is N.

The point-to-point channel capacity is

$$C_{AWGN} = \log_2\left(1+\frac{S}{N}\right), \tag{41}$$

where the signal power is S and the noise power is N.

From Eq. (31), the degree of freedom is [14].

$$DoF = \lim_{x\to\infty}\left(\frac{1+\frac{S}{N}}{1+\frac{S}{N}}\right) = 1, \tag{42}$$

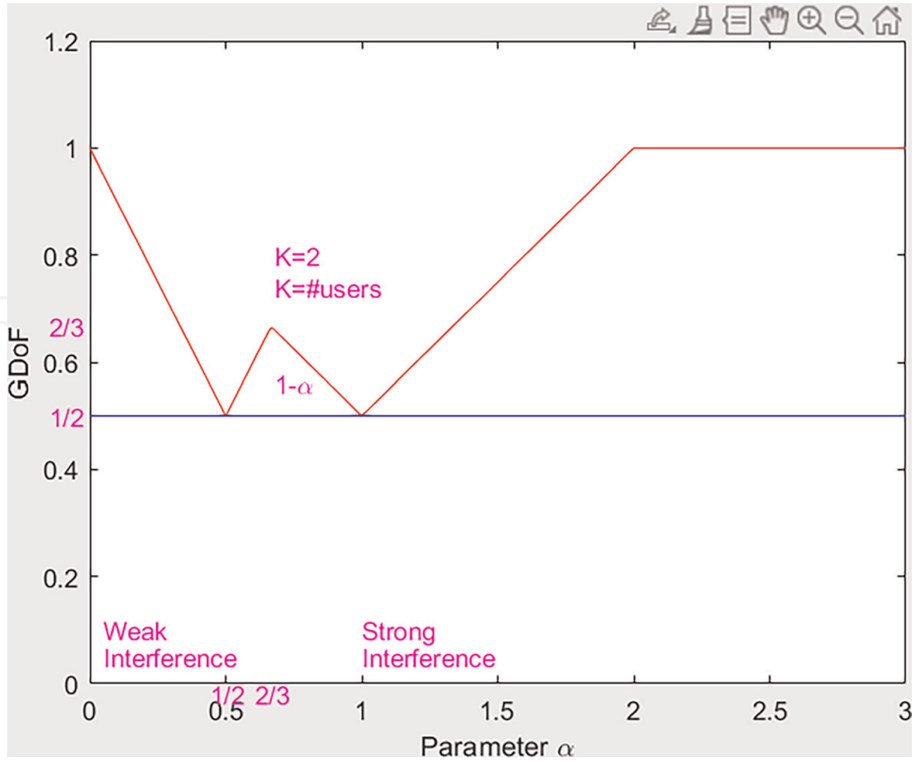

Figure 5.
Generalized degree of freedom for Gaussian Channel (W curve).

And the achievable rate is orthogonalized as

$$\sum_{i=1}^{K} R_i = \log_2\left(1+\frac{\sum_{i=1}^{K} P_i}{N}\right), \tag{43}$$

where K means the number of users.
For two users,

$$2R = \log_2\left(1+2\frac{P}{N}\right) = \log_2(1+2SNR). \tag{44}$$

Therefore, the achievable rate is,

$$R = \frac{1}{2}\log_2(1+2SNR). \tag{45}$$

SNR = 19 and SNR = 31 case:

$$\text{The capacity}: C = \frac{1}{2}\log_2\left(1+\frac{19}{31}\right) = \frac{1}{2}\log_2(1+0.61) = 0.34 \tag{46}$$

$$\text{Achievable rate}: \begin{aligned} 2R &= \log_2\left(1+2\left(\frac{19}{31}\right)\right) \\ 2R &= \log_2(2.22) \\ 2R &= 1.15 \\ R &= 0.57 \end{aligned} \tag{47}$$

And the degree of freedom,

$$DoF = \lim_{SNR\to\infty}\left(\frac{R}{\log_2(2SNR)}\right) \approx \frac{1}{2}\left(\frac{\log_2(1+2SNR)}{\log_2(2SNR)}\right) \approx \frac{1}{2}. \tag{48}$$

On the condition that the ratio $\alpha = \frac{\log_2 INR}{\log_2 SNR}$ is fixed and the strength of the signal is much larger than that of interference and noise, it is able to treat interference as noise. Therefore, the achievable rate is represented by

$$R = \log_2\left(1+\frac{SNR}{1+INR}\right). \tag{49}$$

From Eq. (49), the *DoF* is represented by [14].

$$\begin{aligned} DoF &= \lim_{SNR\to\infty}\left(\frac{R}{\log_2\left(\frac{SNR}{1+INR}\right)}\right) = \left(\frac{\log_2\left(\frac{SNR}{1+INR}\right)}{\log_2(SNR)}\right) \approx \left(\frac{\log_2\left(\frac{SNR}{INR}\right)}{\log_2(SNR)}\right) \\ &= \left(\frac{\log_2(SNR)-\log_2(INR)}{\log_2(SNR)}\right) = 1-\left(\frac{\log_2(INR)}{\log_2(SNR)}\right) = (1-\alpha). \end{aligned} \tag{50}$$

In the conventional binary symmetric channel, p is a random variable and a large amount of resources are used up to make an estimate of p corresponding to the given channel. By the way, p can be determined deterministically for the RNA base matrix $[C\ U; A\ G]$, which is either 0.38 or 0.62. Because the specific value of p is given, the channel estimation should be investigated. The reason why the specific numerical values are selected is that for the RNA model, its maximum channel capacity is maintained even if p is determined deterministically, the variance of signal is not large, and a generalized *DoF*'s point of view shows a reasonable performance in the W curve. In the actual implementation, the receiver has to be satisfied with the 1-α = p shown in **Figure 2**. Under this circumstance, signal strength and the interference intensity are important to analyze the given channel where strong interference environment and weak interference environment are classified according to α. To take an example, if α = 1-p = 0.38, we need to analyze the strong interference channel. If α = 1-p = 0.62, we need to analyze the weak interference channel. This p estimation is able to minimize performance degradation in the binary symmetric channel while significantly reducing computational complexity. The *GDoF* curve of two user interference symmetric channel in **Figure 5** is the highly recognizable "W" curve shown that it greatly improves understanding of interference channel by identifying two regimes. From the abovementioned example, over the symmetric channel, when α = 0.62, the signal is relatively stronger than interference. By the way, when α = 0.38, signal is relatively weaker than interference.

6. RNA genetic code constructed by block circulant jacket matrix

A block circulant Jacket matrix (BCJM) is defined by [7, 12, 13, 15].

$$\mathbf{C}_4 = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{C}_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & a & -a \\ 1 & -1 & -1/a & -1/a \\ a & -a & 1 & 1 \\ -1/a & -1/a & 1 & -1 \end{pmatrix}_{a=1} \tag{51}$$
$$= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix} = \left(\begin{array}{cc:cc} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ \hdashline 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \end{array}\right),$$

where $\mathbf{C}_0$ and $\mathbf{C}_1$ are the Hadamard matrix.

The circulant submatrices are 2×2 matrices, whose entries are moved by block diagonal cyclic shifts. These submatrices are block circulant Jacket matrices. The BCJM $\mathbf{C}_4$ is defined by

$$\underbrace{\mathbf{C}_4 \triangleq \mathbf{I}_0 \otimes \mathbf{C}_0' + \mathbf{I}_1 \otimes \mathbf{C}_1}, \tag{52}$$

where $\mathbf{I}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{I}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{C}_0' = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and $\mathbf{C}_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, while $\otimes$ is the Kronecker product.

From Eq. (52), the genetic matrix $[C\ U; A\ G]^3$ generates RNA sequences such as [12, 13].

$$\mathbf{P}^1 = \begin{pmatrix} C & U \\ A & G \end{pmatrix}, \ \mathbf{P}^2 = \begin{pmatrix} C & U \\ A & G \end{pmatrix} \otimes \begin{pmatrix} C & U \\ A & G \end{pmatrix}, \ \mathbf{P}^3 = \begin{pmatrix} C & U \\ A & G \end{pmatrix}^2 \otimes \begin{pmatrix} C & U \\ A & G \end{pmatrix}, \tag{53}$$

where $\otimes$ denotes the Kronecker product. RNA consists of the sequence of 4 bases where *C, U*, *A,* and *G* indicate cytosine, uracil, adenine, and guanine, on an individual basis.

According to the theory of noise-immunity coding, for 64 triplets, by comparing them with strong roots and weak roots, it is able to construct a mosaic gene matrix $[C\ U; A\ G]^3$. If any triplet belongs to one of the strong roots, it is substituted for 1. In an analogous fashion, if any triplet is included with one of the weak roots, it is replaced with -1. Here, the strong roots are $(CC, CU, CG, AC, UC, GC, GU, GG)$ and $(CA, AA, AU, AG, UA, UU, UG, GA)$ are the weak roots, which results in the singular Rademacher matrix $\mathbf{R}_8$ is in **Table 3** [6, 16].

A novel encoding scheme is proposed as

$$\mathbf{R}_8 = \left(\begin{array}{cccc:cccc} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \hdashline 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{array}\right). \tag{54}$$

	000 (0)	**001 (1)**	**010 (2)**	**011 (3)**	**100 (4)**	**101 (5)**	**110 (6)**	**111 (7)**
000 (0)	CCC 000	CCU 001	CUC 010	CUU 011	UCC 100	UCU 101	UUC 110	UUU 111
001 (1)	CCA 001	CCG 000	CUA 011	CUG 010	UCA 101	UCG 100	UUA 111	UUG 110
010 (2)	CAC 010	CAU 011	CGC 000	CGU 001	UAC 110	UAU 111	UGC 100	UGU 101
011 (3)	CAA 011	CAG 010	CGA 001	CGG 000	UAA 111	UAG 110	UGA 101	UGG 100
100 (4)	ACC 100	ACU 101	AUC 110	AUU 111	GCC 000	GCU 001	GUC 010	GUU 011
101 (5)	ACA 101	ACG 100	AUA 111	AUG 110	GCA 001	GCG 000	GUA 011	GUG 010
110 (6)	AAC 110	AAU 111	AGC 100	AGU 101	GAC 010	GAU 011	GGC 000	GGU 001
111 (7)	AAA 111	AAG 110	AGA 101	AGG 100	GAA 011	GAG 010	GGA 001	GGG 000

Table 3.
[C U;A G]³ code [6, 16].

The Eq. (54) gives a hint of the DNA double helix.
Make a mental note to ensure that

$$\underbrace{\mathbf{R}_8 \triangleq \mathbf{I}_0 \otimes \mathbf{C}_0 \otimes \mathbf{P}_2 + \mathbf{I}_1 \otimes \mathbf{C}_1 \otimes \mathbf{P}_2}, \tag{55}$$

where $\mathbf{I}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{I}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{C}_0 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \mathbf{C}_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, and $\mathbf{P}_2$ is the double stochastic permutation matrix represented by $\mathbf{P}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Eq. (54) has a series of redundant rows which just repeat and are able to be canceled. From the Rademacher matrix $\mathbf{R}_8$, one version of its mosaic gene matrices can be reached as

$$\mathbf{R}_8' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}. \tag{56}$$

Furthermore, by canceling the repeated column from Eq. (56) by means of CRISPR, another version of the mosaic gene matrices can be reached as Eq. (57), which is a singular RNA matrix.

$$\mathbf{R}_4'' = \left(\begin{array}{cc:cc} 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ \hdashline 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{array}\right) = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}, \tag{57}$$

where $\mathbf{C}_0 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ and $\mathbf{C}_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$. These matrices are able to be expanded into the DNA double helix or the RNA single strand, which indicates the process by that DNA replicates its genetic information for itself, which is transcribed into RNA and used to synthesize protein for its translation. Therefore,

$$\underbrace{\mathbf{R}_4'' \triangleq \mathbf{I}_0 \otimes \mathbf{C}_0 + \mathbf{I}_1 \otimes \mathbf{C}_1}, \tag{58}$$

where $\mathbf{C}_0$ has eigenvalues such that $\lambda_1^{(1)} = 1 + i$ and $\lambda_2^{(1)} = 1 - i$, and their eigenvectors $\varsigma_1 = (1 \quad -i)^T$ and $\varsigma_2 = (1 \quad i)^T$, correspondingly. In addition, $\mathbf{C}_1$ has eigenvalues such that $\lambda_1^{(2)} = \sqrt{2}$ and $\lambda_2^{(2)} = -\sqrt{2}$ where their eigenvectors $\varsigma_1 = \left(-1 + \sqrt{2} \quad 1\right)^T$ and $\varsigma_1 = \left(-1 - \sqrt{2} \quad 1\right)^T$ on an individual basis [3, 17]. Then,

$$\mathbf{R}_4'' \otimes \mathbf{P}_2 \Rightarrow \mathbf{R}_8 = \mathbf{R}_{4\times 2^k}, \tag{59}$$

where k = 1.

7. Symmetric genetic jacket block matrix

It is demonstrated that the genomatrices are constructed based on the kernel $[C\,A; U\,G]$ and the mosaic genomatrices $[C\,A; U\,G]^3$ are built by a series of Kronecker products, which are expanded by permuting the 4 bases C, A, U, and G on their locations in the matrix.

7.1 Permutation scheme from upper to lower

Following this scheme, we are confronted with 24 variants of genomatrices, which distinguish them from each other by replacing their subsets by the kernel $[C\,A; U\,G]$. To take an analogous instance, by applying the upper-low scheme to $[C\,A; U\,G]$, the standard genetic code is expanded into $[U \quad C \quad A \quad G]^T \otimes [U \quad C \quad A \quad G] \otimes [U \quad C \quad A \quad G]^T$, where ^{T} is the transpose. Analogous to Eq. (56), one version of variants of genomatrices is constructed as

$$\left[\begin{array}{cccc:cccc} -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hdashline -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{array}\right]^T \left[\begin{array}{cccc:cccc} -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hdashline -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{array}\right]$$

$$= \left(\underbrace{[1\,0] \otimes \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right)}_{} + \underbrace{[0\,1] \otimes \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \right)}_{} \right) \otimes \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{60}$$

Eq. (60) is also another version of variants of genomatrices by a series of Kronecker product on $[1\ 1\ 1\ 1]^T$, which is expanded into Eq. (61) indicating the process transcribing from $\mathbf{R}_8$ DNA to $\mathbf{R}_4''$ RNA.

$$\mathbf{R}_4'' = \left[\begin{array}{cc:cc} -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \\ \hdashline -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \end{array}\right] = \underbrace{[1\ 0]\otimes\left(\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}\right)}+\underbrace{[0\ 1]\otimes\left(\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & -1\\ -1 & 1\end{bmatrix}\right)}. \tag{61}$$

Example 7.1. If $A = U$, $C = G$, we are confronted with six versions of variants of the genomatrices constructed by a series of Kronecker product of the kernel $[C\ A; U\ G]$.

$$\begin{aligned}\begin{bmatrix} A & C \\ U & G \end{bmatrix} &= \begin{bmatrix} -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \\ &= [1\ \ 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ -1 & -1\end{bmatrix}+[0\ \ 1]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix},\end{aligned} \tag{62}$$

which is expanded into Eq. (63) and Eq. (64). These are other versions of variants of genomatrices.

$$\begin{aligned}\begin{bmatrix} A & G \\ U & C \end{bmatrix} &= \begin{bmatrix} -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix}, \\ &= [1\ \ 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & -1\\ -1 & 1\end{bmatrix}+[0\ \ 1]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}\end{aligned} \tag{63}$$

$$\begin{aligned}\begin{bmatrix} G & U \\ C & A \end{bmatrix} &= \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \\ &= [1\ \ 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}+[0\ \ 1]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & -1\\ 1 & -1\end{bmatrix}\end{aligned} \tag{64}$$

$$\begin{aligned}\begin{bmatrix} C & U \\ G & A \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}, \\ &= [1\ \ 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}+[0\ \ 1]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}1 & -1\\ -1 & -1\end{bmatrix}\end{aligned} \tag{65}$$

$$\begin{bmatrix} C & A \\ G & U \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}, \quad (66)$$

$$= [1 \quad 0] \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} + [0 \quad 1] \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} G & A \\ C & U \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \end{bmatrix}. \quad (67)$$

$$= [1 \quad 0] \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} + [0 \quad 1] \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$$

Eq. (62–67) are six versions of variants of genomatrices, which indicate six half pairs expanded from symmetric RNA genetic matrices by an upper-lower scheme. In other words, they are constructed by rotating the block in the direction from upper to low or vice versa.

7.2 Permutation scheme from left to right

Following this scheme, we are confronted with 6 variants of genomatrices, which distinguish them from each other with the kernel $[C\,A; U\,G]$. To take an analogous instance, by applying the left-right scheme to $[C\,A; U\,G]$, the standard genetic code is expanded into $\mathbf{R}_8$

$$\left(\begin{array}{cccc:cccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ \hdashline 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{array}\right) \quad (68)$$

$$= \left[\begin{array}{l} \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes (1 \quad 1) \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}} \\ + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes (1 \quad 1) \otimes \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}} \end{array}\right] \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Eq. (68) is also another version of variants of genomatrices by a series of Kronecker product on [1 1;1 1], which is expanded into Eq. (69) indicating the process transcribing from $\mathbf{R}_8$ DNA to $\mathbf{R}_4''$ RNA.

$$\mathbf{R}_4'' = \left[\begin{array}{cc:cc} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ \hdashline +1 & -1 & +1 & -1 \\ -1 & -1 & -1 & -1 \end{array}\right] = \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}}_{}. \tag{69}$$

Example 7.2. If $A = U$, $C = G$, we are confronted with six versions of variants of the genomatrices constructed by a series of Kronecker product of the kernel $[C\ A; U\ G]$.

$$\begin{pmatrix} C & G \\ U & A \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}$$

$$= \left[\underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}}_{}\right], \tag{70}$$

which is expanded into Eq. (71) and Eq. (72). These are other versions of variants of genomatrices.

$$\begin{bmatrix} G & C \\ U & A \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 \end{bmatrix}, \tag{71}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} U & A \\ C & G \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \tag{72}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} A & U \\ G & C \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad (73)$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} G & C \\ A & U \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 \end{bmatrix}, \quad (74)$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} C & G \\ A & U \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}. \quad (75)$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$$

Eqs. (70)-(75) are 6 versions of variants of genomatrices, which indicate six half pairs expanded from symmetric RNA genetic matrices by the left-right scheme. In other words, they are constructed by rotating the block in the direction from upper to low or vice versa.

7.3 Block Circulant jacket matrix

Construct a block matrix $[C]_N$ by Jacket matrices $[C_0]_p$ and $[C_1]_p$ such as $[C]_N = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}$ where its order N is $2p$. This matrix is called block circulant if only if $C_0 C_1^{RT} + C_1^{RT} C_0 = [0]_N,$ where ^{RT} is the reciprocal transpose. In other words, $[C]_N$ is a block circulant Jacket matrix (BCJM) [12, 13, 15, 18]. From the fact that $C_0 C_0^{RT} = p[I]_p$ and $C_1 C_1^{RT} = p[I]_p,$ C_0 and C_1 are Jacket matrices. Look back on the fact that $[C]_N$ is a Jacket matrix if only if $[C][C]^{RT} = NI_N,$ where ^{RT} is the reciprocal transpose. Therefore, $[C]$ is a Jacket matrix if only if

$$[C][C]^{RT} = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix} \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}^{RT} = \begin{pmatrix} 2p[I]_p & C_0 C_1^{RT} + C_1^{RT} C_0 \\ C_0 C_1^{RT} + C_1^{RT} C_0 & 2p[I]_p \end{pmatrix} = NI_N, \quad (76)$$

where ^{RT} is the reciprocal transpose. Therefore, Eq. (76) results in plenty of BCJMs.

Example 7.3. Two 2×2 matrices are given such as

$$C_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, C_1 = \begin{pmatrix} a & -a \\ -1/a & -1/a \end{pmatrix}.$$

It is easy to know that $C_0 C_0^{RT} = 2[I]_2$ and $C_1 C_1^{RT} = 2[I]_2$ are satisfied. Therefore, C_0 and C_1 are Jacket matrices.

Moreover,

$$C_0 C_1^{RT} + C_1^{RT} C_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} -1/a & -a \\ -1/a & -a \end{pmatrix} + \begin{pmatrix} a & -a \\ -1/a & -1/a \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = [0]_2. \quad (77)$$

8. General pattern of block circulant symmetric genetic jacket matrix

We present 24(=4 × ${}_4\mathbf{C}_2$) DNA classes of genomatrices with their own characteristics. The main kernel of Eq. (78) is

$$\underbrace{E}_{\text{Position}} \otimes \underbrace{\{(I_0 \otimes A) + (I_1 \otimes B)\}}_{\text{Main Body Kernel}} \otimes \underbrace{F}_{\text{Extending}}. \quad (78)$$

Eq. (58) is an RNA pattern by the main kernel. By applying an upper-lower or left-right scheme to the genetic matrix, the position matrix **E** creates the patterns analogous to Eq. (61, 69). Analogously, by applying the upper-lower and left-right scheme to the genetic matrix, the extending matrix F creates the patterns analogous to Eq. (60, 68).

South Korea's national flag stands for different symbols of trigrams and Yin-Yang located in its middle, which is analogous to that of **Figure 6**. We present 24 versions of variants of genomatrices, which distinguish from each other by replacing their subsets with the kernel shown in **Figure 6** like its left-hand side $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1]$, its right-hand

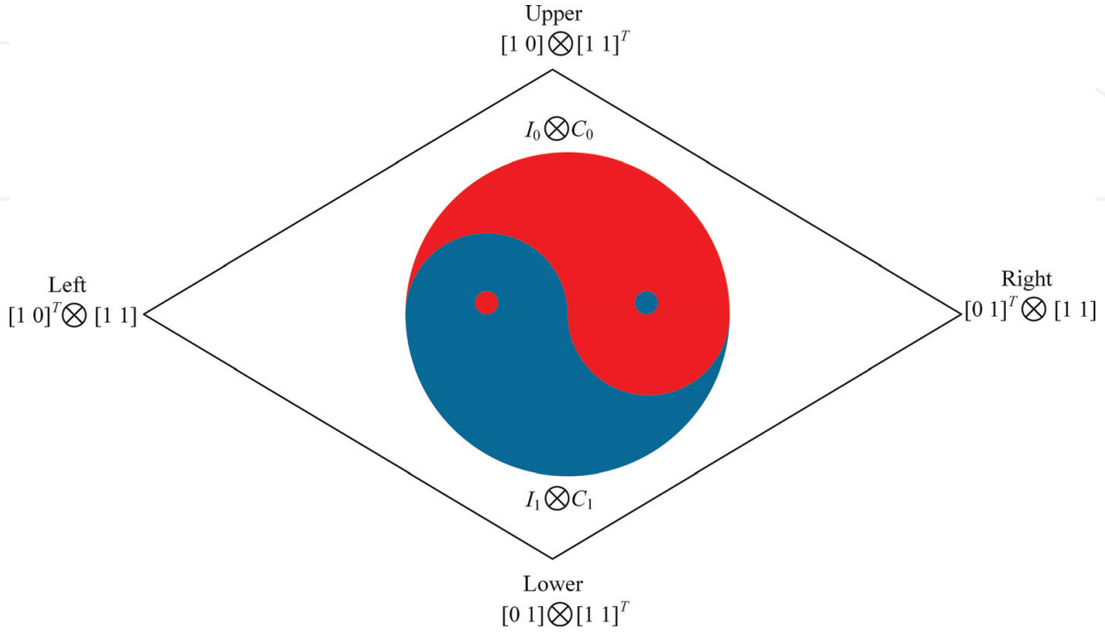

Figure 6.
General pattern by block circulant, upper-lower, and left–right scheme: Normal case.

side $\begin{bmatrix}0\\1\end{bmatrix}\otimes[1\quad 1]$, its upper position $(1\quad 0)\otimes\begin{pmatrix}1\\1\end{pmatrix}$, its lower position $(0\quad 1)\otimes\begin{pmatrix}1\\1\end{pmatrix}$, and its center part $I_0\otimes C_0+I_1\otimes C_1$, on an individual basis.

From the fact that $(1\quad 0)\otimes\begin{pmatrix}1\\1\end{pmatrix}\leftrightarrow(0\quad 1)\otimes\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\0\end{pmatrix}\otimes(1\quad 1)\leftrightarrow\begin{pmatrix}0\\1\end{pmatrix}\otimes(1\quad 1)$, upper symmetric genetic matrices are complementary with lower ones while left ones are complementary with right ones.

In addition, the pattern is created by block circulant, upper-lower, and left–right scheme on the ½ symmetric block, which are analyzed in three cases.

Case 1. Block circulant scheme

$$\begin{bmatrix}C & U\\ A & G\end{bmatrix}=\begin{bmatrix}1 & 1 & 1 & -1\\ -1 & 1 & -1 & -1\\ 1 & -1 & 1 & 1\\ -1 & -1 & 1 & -1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & A^{diag}\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}+\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\otimes\begin{bmatrix}1 & -1\\ -1 & -1\end{bmatrix}. \tag{79}$$

$$\begin{bmatrix}U & C\\ G & A\end{bmatrix}=\begin{bmatrix}-1 & 1 & 1 & 1\\ -1 & -1 & 1 & -1\\ 1 & 1 & -1 & 1\\ -1 & 1 & -1 & -1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ -1 & -1\end{bmatrix}+\begin{bmatrix}0 & 1\\ A^{Anti-diag} & 0\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}. \tag{80}$$

Case 2. Upper-lower scheme

$$\begin{bmatrix}U & G\\ A & C\end{bmatrix}=\begin{bmatrix}-1 & -1 & 1 & 1\\ -1 & 1 & -1 & 1\\ -1 & -1 & 1 & -1\\ -1 & 1 & 1 & 1\end{bmatrix}=[1\quad 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & -1\\ -1 & 1\end{bmatrix}+[0\quad A^{Upper}]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}. \tag{81}$$

$$\begin{bmatrix}U & C\\ A & G\end{bmatrix}=\begin{bmatrix}-1 & 1 & 1 & 1\\ -1 & -1 & 1 & -1\\ -1 & 1 & 1 & 1\\ -1 & -1 & -1 & 1\end{bmatrix}=[1\quad 0]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}-1 & 1\\ -1 & -1\end{bmatrix}+[0\quad A^{Lower}]\otimes\begin{bmatrix}1\\1\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}. \tag{82}$$

Case 3. Left-right scheme

$$\begin{bmatrix} A & U \\ C & G \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ A^{Left} \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{83}$$

$$\begin{bmatrix} U & A \\ G & C \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ A^{Right} \end{bmatrix} \otimes [1 \quad 1] \otimes \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \tag{84}$$

Eq. (79) is a block circulant while Eq. (80) is not. Meanwhile, one part of Eq. (81, 82) is upper-lower symmetric while the other is not. By the way, one part of Eq. (83, 84) is left–right symmetric while the other part is not. **Figure 7** shows a certain pattern constructed by a series of the product of $[C\ A; U\ G]$ as well as a distorted pattern in comparison with that in **Figure 6**. Therefore, these are called sickness pattern, which can cover COVID-19.

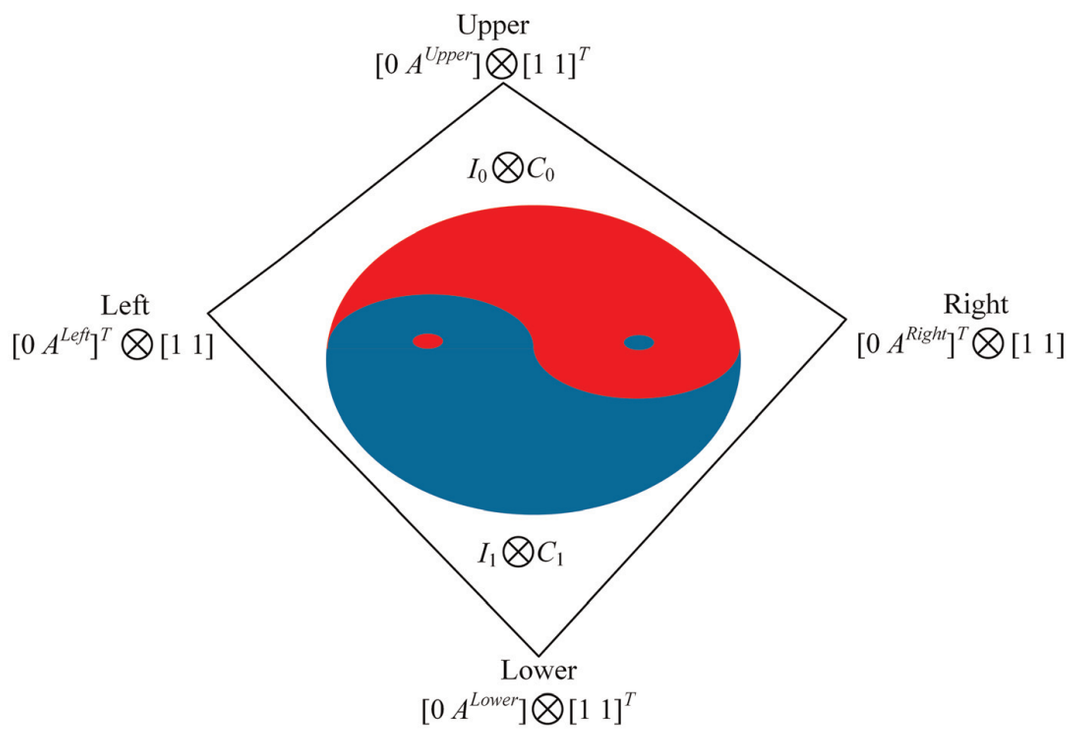

Figure 7.
Abnormal pattern by block circulant, upper-lower, and left–right scheme.

To take an analogous instance,

$$\begin{pmatrix} C & U \\ A & G \end{pmatrix} \Rightarrow \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{85}$$

Make a mental note to ensure.
Case 1. $A \neq D,\ B = C$ and $A = D,\ B \neq C$.
Case 2. $A = C,\ B \neq D$ and $A \neq C,\ B = D$.
Case 3: $A = B,\ C \neq D$ and $A \neq B,\ C = D$.
From the aforementioned processes, we are confronted with six half symmetric blocks such as $\begin{pmatrix} C & U \\ A & G \end{pmatrix}, \begin{pmatrix} U & C \\ G & A \end{pmatrix}, \begin{pmatrix} U & G \\ A & C \end{pmatrix}, \begin{pmatrix} U & C \\ A & G \end{pmatrix}, \begin{pmatrix} A & U \\ C & G \end{pmatrix}$, and $\begin{pmatrix} U & A \\ G & C \end{pmatrix}$.

9. Conclusion

We show the experimental results of $C = G = 19\%$ and $A = U = T = 31\%$ for the COVID-19 with the RNA base matrix $[C\ U; A\ G]$, which are expanded into our mathematical proof based on the information theory of doubly stochastic matrix. RNA entropy cannot reach the Shannon entropy because the probabilities of its bases are 23% away from a half that is exactly identical to the sum of its variances. In other words, there is a difference between Shannon capacity and RNA capacity, which is identical to the sum of variances of RNA base random variables because they are unable to become a half over a symmetric channel. We present a straightforward way of laying out a mathematical basis for double helix DNA in the process of reverse transcription from RNA to DNA, which is straightforward and explicit by decomposing a DNA matrix into sparse matrices which have non-redundant columns and rows. And we introduce a general pattern by block circulant, upper-lower, and left–right scheme, which is applied to the correct communication as well as means the healthy condition because it perfectly consists of 4 bases. Furthermore, we introduce an abnormal pattern by block circulant, upper-lower, and left–right scheme, which covers the distorted signal as well as COVID-19. The Equation 57, RNA matrix is the same as the Reference 11 USA patent MIMO Comm. definition 3.1 matrix.

Conflict of interest

The authors declare no conflict of interest.

Author details

Sung Kook Lee[1] and Moon Ho Lee[2*]

1 Jeju International School, Jeju, Republic of Korea

2 Department of Electronics, Jeonbuk National University, Jeonju, Republic of Korea

*Address all correspondence to: moonho@jbnu.ac.kr

IntechOpen

References

[1] Chargaff E, Zamenhof S, Green C. Human desoxypentose nucleic acid: Composition of human desoxypentose nucleic acid. Nature. 1950;**165**:756-757. DOI: 10.1038/165756b0

[2] Watson J, Crick F. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 1953; **171**:737-738. DOI: 10.1038/171737a0

[3] Temin HM. Nature of the provirus of Rous sarcoma. National Cancer Institute Monograph. 1964;**17**:557-570

[4] Lee MH, Lee SK, Cho KM. A Life Ecosystem Management With DNA Base Complementarity. Moscow: Proceedings of the International Conference of Artificial Intelligence, Medical Engineering, Education (AIMEE 2018); 6–8 October 2018; Springer Nature; 2020

[5] Papoulis A, Pillai SU. Probability, Random Variables and Stochastic Process. 4th ed. Boston: McGraw Hill; 2002

[6] He M, Petoukhov S. Mathematics of Bioinformatics: Theory, Practice, and Applications. 1st ed. New Jersey: John Wiley & Sons; 2010. DOI: 10.1002/ 9780470904640

[7] Lee SK, Park DC, Lee MH. RNA genetic 8 by 8 matrix construction from the block circulant Jacket matrix. Springer Nature: Proceedings of Symmetry Festival 2016; 18-22 July 2016, Vienna, Cham; 2017

[8] Shannon CE. A Mathematical Theory of Communication. The Bell System Technical Journal. 1948;**27**:31-423-623-656. DOI: 10.1002/j.1538-7305.1948. tb01338.x\

[9] Azgari C, Kilinc Z, Turhan B, Circi D, Adebali O. The mutation profile of SARS-CoV-2 is primarily shaped by the host antiviral defense. Viruses. 2021; **13**(3):394. DOI: 10.3390/v13030394

[10] Berkhout B, Hemert VF. On the biased nucleotide composition of the human coronavirus RNA genome. Virus Research. 2015;**202**:41-47. DOI: 10.1016/ j.virusres.2014.11.031

[11] Xia X. Extreme Genomic CpG Deficiency in SARS-CoV-2 and Evasion of Host Antiviral Defense. Molecular Biology and Evolution. 2020;**37**(9):2699-2705. DOI: 10.1093/molbev/msaa094

[12] Lee MH, Hai H, Zhang XD. MIMO Communication Method and System using the Block Circulant Jacket Matrix. United States Patent US 009356671B1 [Internet]. 31 May 2016. Available from: https://patentimages.storage.googleapis.com/cb/46/34/4acf23e5a9b6e1/US9356671.pdf [Accessed: 12 December 2021]

[13] Lee MH. Jacket Matrices: Construction and Its Application for Fast Cooperative Wireless Signal Processing. 1st ed. Germany, Saarbrucken: LAP LAMBERT Academic Publishing; 2012

[14] Tse D, Viswanath P. Fundamentals of Wireless Communication. 1st ed. New York: Cambridge University Press; 2005. DOI: 10.1017/CBO9780511807213

[15] Wikipedia, the free encyclopedia. Jacket Matrix [Internet]. 1999. Available from: https://en.wikipedia.org/wiki/Jacket_matrix [Accessed: 12 December 2021]

[16] Rumer YB. Translation of 'Systematization of Codons in the Genetic Code [II]' by Yu. B. Rumer (1968). Royal Society. 2016;**374**:2063. DOI: 10.1098/rsta.2015.0447

[17] Lee MH, Hai H, Lee SK, Petoukhov SV. A Mathematical Proof of Double Helix DNA to Reverse Transcription RNA for Bioinformatics. Moscow: Proceedings of the 1st International Conference of Artificial Intelligence, Medical Engineering, and Education (AIMEE 2017); 21–23 August 2017; Springer Nature; 2018

[18] Chen Z, Lee MH, Zeng G. Fast cocyclic Jacket transform. IEEE trans. on Signal Processing. 2008;**56**(5):2143-2148. DOI: 10.1109/TSP.2007.912895