

# ГИПЕРКОМПЛЕКСНЫЕ ЧИСЛА И АЛГЕБРАИЧЕСКАЯ СИСТЕМА ГЕНЕТИЧЕСКИХ АЛФАВИТОВ. ЭЛЕМЕНТЫ АЛГЕБРАИЧЕСКОЙ БИОЛОГИИ

С.В. Петухов

*Институт машиноведения РАН, Москва, Россия*  
spetoukhov@gmail.com

Представлены результаты исследования многоуровневой системы взаимосвязанных молекулярно-генетических алфавитов на основе применения матричных методов теории помехоустойчивого кодирования. Эти исследования выявили связи данной системы алфавитов с системами гиперкомплексных чисел (кватернионами Гамильтона и сплит-кватернионами Кокла и их расширениями), кронекеровскими семействами матриц, ортогональными системами функций Радемахера и Уолша, матрицами Адамара и др. Отмечаются структурные параллелизмы между системой молекулярно-генетических алфавитов и системой наследования признаков у целостных организмов, подчиняющейся законам Менделя и представляемой классическими решетками Пеннета. Система молекулярно-генетических алфавитов, общая для всех живых организмов, своими алгебраическими свойствами подсказывает новый — алгебраический — путь познания живой материи и развития алгебраической биологии, связанной с гиперкомплексными числами. Живая материя, обеспечивающая передачу наследственной информации по цепи поколений, предстает информационной сущностью, глубоко алгебраичной по своей природе.

**Ключевые слова:** генетические алфавиты, гиперкомплексные числа, функции Радемахера, функции Уолша, матрицы Адамара, циклические группы, решетки Пеннета, алгебраическая биология.

## 1 Введение

Достижения молекулярной генетики привели к новому пониманию самой жизни: «Жизнь есть партнерство между генами и математикой» [1]. Но какая именно математика является партнером генетического кода?

Пытаясь нащупать такую математику, автор обратился к изучению многоуровневой системы взаимно связанных молекулярно-генетических алфавитов. На этом пути неожиданно обнаружили связи генетической системы с известными формализмами инженерно-математической теории помехоустойчивого кодирования информации: кронекеровским произведением матриц, ортогональными системами функций Радемахера и Уолша, таблицами умножения гиперкомплексных числовых систем, матрицами Адамара и т.д. [2, 3, 4]. Настоящая статья посвящена некоторым результатам такого изучения феноменологической системы взаимно связанных генетических алфавитов.

Алфавиты играют базисную роль в технологиях коммуникации. В любой коммуникационной системе «передатчик-приемник» приемник всегда знает алфавит сигналов, который использует передатчик для сообщений. В лингвистике алфавиты обычно имеют многоуровневую структуру, поскольку содержат подмножества гласных и согласных звуков, причем подмножество гласных состоит из подмножеств долгих и кратких гласных, а подмножество согласных — из подмножеств звонких и глухих согласных и т.д. Количества членов во всех этих частях лингвистических алфавитов не связаны друг с другом какими-либо известными алгебраическими закономерностями. Нами обнаружено, что в многоуровневой системе генетических алфавитов ситуация существенно иная: многие

части этой системы тесно взаимосвязаны посредством содержательных алгебраических закономерностей и формализмов, которые широко известны в математике и технике связи.

Молекулярно-генетическая система живого вещества содержит следующие алфавиты, каждый из которых может быть рассмотрен как отдельный алфавит или как часть сложной алфавитной системы:

- 4-буквенный алфавит азотистых оснований;
- 64-буквенный алфавит триплетов;
- 2-буквенный алфавит «слабых» и «сильных корней» триплетов;
- 20-буквенный алфавит аминокислот;
- 2-буквенный алфавит пуринов и пиримидинов;
- 2-буквенный алфавит сильных и слабых водородных связей;
- 2-буквенный алфавит «кетто и амина»,

и т.д. (см. расширенный список генетических алфавитов в статье [5]).

С учетом этого молекулярно-генетическая система является полиязычной. Любая последовательность азотистых оснований в молекулах наследственности ДНК и РНК может читаться как полиязычное послание, интерпретируемое в зависимости от настройки считывающего устройства на тот или иной алфавит. Можно добавить, что распространенное выражение «генетический код» означает взаимное соответствие между элементами двух из перечисленных алфавитов: алфавита триплетов и алфавита аминокислот и стоп-кодона.

Генетическая информация, записанная на уровне микромира генетических молекул, диктует конструкции в макромире живых организмов в условиях сильных шумов и помех. Закон Менделя независимого наследования различительных признаков (например, цвета волос, кожи и глаз наследуются независимо друг от друга) свидетельствует о том, что этот диктат реализуется через различные независимые каналы передачи информации посредством неизвестных алгоритмов многоканального помехоустойчивого кодирования. Соответственно каждый организм предстает алгоритмической машиной многоканального помехоустойчивого кодирования. Для познания этой машины следует использовать теорию помехоустойчивого кодирования.

## 2 Матричные представления генетических алфавитов

Теория помехоустойчивого кодирования базируется на матричных методах. Например, матричные методы позволяют передавать качественные фотографии поверхности Марса на Землю через миллионы километров сильнейших помех. Для этих целей используются, в частности, кронекеровские семейства матриц Адамара (Рис. 1).

На Рис. 1 символ  $(n)$  означает кронекеровскую (или тензорную) целую степень. Кронекеровское произведение матриц является обычной операцией в технике обработки сигналов, теоретической физике и пр. Оно используется для перехода от пространств с меньшей размерностью к ассоциированным пространствам более высокой размерности. По аналогии с приемами теории связи мы представляем 4-буквенный алфавит азотистых оснований А, С, G, Т/U (аденин, цитозин, гуанин, тимин/урацил) в форме квадратной матрицы  $[C\ T; A\ G]$  (Рис. 2). Этот 4-буквенный алфавит связан с 16-буквенным алфавитом генетических дуплетов и 64-буквенным алфавитом генетических триплетов посредством второй и третьей кронекеровской степени этой матрицы  $[C\ T; A\ G]^{(2)}$  и  $[C\ T; A\ G]^{(3)}$ , которые дают  $(4 * 4)$ - и  $(8 * 8)$ -матрицы с упорядоченным расположением дуплетов и триплетов в них (Рис. 2).

$$H_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}; H_2^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}; H_2^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix}$$

Рис. 1: Начальные матрицы Адамара из семейства  $(2^n * 2^n)$ -матриц Адамара, основанного на кронекеровских степенях  $(n)$  базовой матрицы  $H_2$  этого семейства.

$$[C T; A G] = \begin{bmatrix} C & T \\ A & G \end{bmatrix}; [C T; A G]^{(2)} = \begin{bmatrix} CC & CT & TC & TT \\ CA & CG & TA & TG \\ AC & AT & GC & GT \\ AA & AG & GA & GG \end{bmatrix}$$

$$[C T; A G]^{(3)} = \begin{bmatrix} CCC & CCT & CTC & CTT & TCC & TCT & TTC & TTT \\ CCA & CCG & CTA & CTG & TCA & TCG & TTA & TTG \\ CAC & CAT & CGC & CGT & TAC & TAG & TGC & TGT \\ CAA & CAG & CGA & CGG & TAA & TAG & TGA & TGG \\ ACC & ACT & ATC & ATT & GCC & GCT & GTC & GTT \\ ACA & ACG & ATA & ATG & GCA & GCG & GTA & GTG \\ AAC & AAT & AGC & AGT & GAC & GAT & GGC & GGT \\ AAA & AAG & AGA & AGG & GAA & GAG & GGA & GGG \end{bmatrix}$$

Рис. 2: Генетическая  $(2 * 2)$ -матрица  $[C T; A G]$  азотистых оснований и геноматрицы ее второй и третьей кронекеровских степеней, содержащие упорядоченные множества 16 дуплетов и 64 триплетов. Черным выделены триплеты с сильными корнями и дуплеты, соответствующие таким корням (пояснение в тексте).

Обратимся теперь к алфавиту сильных и слабых корней триплетов. Две первые позиции каждого триплета называются его «корнем». Если все 4 триплета с одинаковым корнем кодируют одну и ту же аминокислоту, то такой корень называется «сильным» (например, CGC, CGA, CGT, CGG кодируют одну аминокислоту Arg). В противном случае корень называется «слабым» [6]. Природа почему-то поделила множество 64 триплетов на два равных подмножества по 32 триплета в каждом: первое подмножество содержит 32 триплета с сильными корнями (CC, CT, CG, AC, TC, GC, GT, GG), а второе — 32 триплета со слабыми корнями (CA, AA, AT, AG, TA, TT, TG, GA) (см. Таблица 1).

Как этот 2-буквенный алфавит сильных и слабых корней сопряжен с 64-буквенным алфавитом триплетов? Существует ли, например, симметрия в расположении триплетов с сильными и слабыми корнями в матрице триплетов  $[C T; A G]^{(3)}$ , которая была построена совершенно формально без упоминания аминокислот и пр.? Заметим, что имеется огромное количество  $64! \approx 10^{89}$  вариантов расположения 64 триплетов в  $(8 \times 8)$ -матрице. Для сравнения, физика оценивает все время существования Вселенной в  $10^{17}$  секунд. Очевидно,

Таблица 1: Подмножества триплетов с сильными (слева) и слабыми корнями в «универсальном» генетическом коде и коде митохондрий человека и позвоночных. Исходные данные взяты с сайта Национального биоинформационного общества США (<http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>).

УНИВЕРСАЛЬНЫЙ (СТАНДАРТНЫЙ) ГЕНЕТИЧЕСКИЙ КОД	
8 подсемейств триплетов с сильными корнями и аминокислоты, кодируемые ими	8 подсемейств триплетов со слабыми корнями и кодируемые аминокислоты
■■■■, ■■■■, ■■■■, ■■■■ → Pro	<u>CAC</u> , <u>CAT</u> , <u>CAA</u> , <u>CAG</u> → His, His, Gln, Gln
■■■■, ■■■■, ■■■■, ■■■■ → Leu	<u>AAC</u> , <u>AAT</u> , <u>AAA</u> , <u>AAG</u> → Asn, Asn, Lys, Lys
■■■■, ■■■■, ■■■■, ■■■■ → Arg	<u>ATC</u> , <u>ATT</u> , <u>ATA</u> , <u>ATG</u> → Ile, Ile, Ile, Met
■■■■, ■■■■, ■■■■, ■■■■ → Thr	<u>AGC</u> , <u>AGT</u> , <u>AGA</u> , <u>AGG</u> → Ser, Ser, Arg, Arg
■■■■, ■■■■, ■■■■, ■■■■ → Ser	<u>TAC</u> , <u>TAT</u> , <u>TAA</u> , <u>TAG</u> → Tyr, Tyr, Stop, Stop
■■■■, ■■■■, ■■■■, ■■■■ → Ala	<u>TTC</u> , <u>TTT</u> , <u>TTA</u> , <u>TTG</u> → Phe, Phe, Leu, Leu
■■■■, ■■■■, ■■■■, ■■■■ → Val	<u>TGC</u> , <u>TGT</u> , <u>TGA</u> , <u>TGG</u> → Cys, Cys, Stop, Trp
■■■■, ■■■■, ■■■■, ■■■■ → Gly	<u>GAC</u> , <u>GAT</u> , <u>GAA</u> , <u>GAG</u> → Asp, Asp, Glu, Glu
ГЕНЕТИЧЕСКИЙ КОД МИТОХОНДРИЙ ПОЗВОНОЧНЫХ	
■■■■, ■■■■, ■■■■, ■■■■ → Pro	<u>CAC</u> , <u>CAT</u> , <u>CAA</u> , <u>CAG</u> → His, His, Gln, Gln
■■■■, ■■■■, ■■■■, ■■■■ → Leu	<u>AAC</u> , <u>AAT</u> , <u>AAA</u> , <u>AAG</u> → Asn, Asn, Lys, Lys
■■■■, ■■■■, ■■■■, ■■■■ → Arg	<u>ATC</u> , <u>ATT</u> , <u>ATA</u> , <u>ATG</u> → Ile, Ile, Met, Met
■■■■, ■■■■, ■■■■, ■■■■ → Thr	<u>AGC</u> , <u>AGT</u> , <u>AGA</u> , <u>AGG</u> → Ser, Ser, Stop, Stop
■■■■, ■■■■, ■■■■, ■■■■ → Ser	<u>TAC</u> , <u>TAT</u> , <u>TAA</u> , <u>TAG</u> → Tyr, Tyr, Stop, Stop
■■■■, ■■■■, ■■■■, ■■■■ → Ala	<u>TTC</u> , <u>TTT</u> , <u>TTA</u> , <u>TTG</u> → Phe, Phe, Leu, Leu
■■■■, ■■■■, ■■■■, ■■■■ → Val	<u>TGC</u> , <u>TGT</u> , <u>TGA</u> , <u>TGG</u> → Cys, Cys, Trp, Trp
■■■■, ■■■■, ■■■■, ■■■■ → Gly	<u>GAC</u> , <u>GAT</u> , <u>GAA</u> , <u>GAG</u> → Asp, Asp, Glu, Glu

что случайное расположение 20 аминокислот и соответствующих триплетов в  $(8 \times 8)$ -матрице почти никогда не даст симметрии в ее половинах, квадрантах и строках. Но неожиданно феноменологическое расположение 32 триплетов с сильными корнями и 32 триплетов со слабыми корнями имеет в этой матрице симметрический характер характер, как и расположение соответствующих сильных и слабых дуплетов в  $(4 * 4)$ -геноматрице  $[C T; A G]^{(2)}$  (Рис. 2):

1. Оба квадранта вдоль каждой диагонали тождественны по мозаике.
2. Верхняя и нижняя половины в каждой матрице зеркально-антисимметричны друг другу по цвету (генетически наследуемое тело каждого человека состоит из двух зеркально-антисимметричных половин, но, как оказывается, подобная зеркальная антисимметрия половин имеет место уже на уровне системы молекулярно-генетических алфавитов).

Какие математические секреты генетической системы отражены в этих мозаичных генетических матрицах взаимно сопряженных генетических алфавитов? Наиболее важным фактом оказывается то, что мозаичный характер каждого столбца в каждой из этих двух геноматриц соответствует нечетной меандровой функции. Такие нечетные меандровые функции хорошо известны в теории обработки сигналов под именем «функций Радемахера». Функции Радемахера содержат только элементы «+1» и «-1». Каждый из столбцов в этих геноматрицах представляет одну из функций Радемахера, если каждый черный (белый) дуплет и триплет в них интерпретировать как элемент +1 (-1).

При этом получают соответствующие радемахеровские представления  $R_4$  матрицы дуплетов  $[C\ T; A\ G]^{(2)}$  и  $R_8$  матрицы триплетов  $[C\ T; A\ G]^{(3)}$ . Рис. 3 показывает эти представления, а также их диадо-сдвиговые декомпозиции на четыре ( $c_0, c_1, c_2, c_3$ ) и восемь ( $r_0, r_1, r_2, r_3, r_4, r_5, r_6, r_7$ ) разреженных матриц соответственно. Матрицы  $c_0$  и  $r_0$  являются единичными.

Автором обнаружено, что каждый из этих наборов из 4 и 8 разреженных матриц замкнут относительно умножения (произведение любых двух матриц из набора дает матрицу из того же набора с точностью до знака) и определяет свою собственную таблицу умножения (Рис. 4).

Таблица умножения на Рис. 4 слева совпадает с таблицей умножения 4-мерной алгебры сплит-кватернионов Дж. Кокла (J. Cockle), открытой им в 1849 году и имеющей ряд приложений в физике [[http://en.wikipedia.org/wiki/Split\\_quaternion](http://en.wikipedia.org/wiki/Split_quaternion)]. Например, она используется в дисковой модели Пуанкаре геометрии Лобачевского. Таким образом, геноматрица дуплетов  $[C\ T; A\ G]^{(2)}$  в ее радемахеровской форме  $R_4$  (Рис. 3) представляет собой сплит-кватернион Кокла с единичными координатами. Радемахеровская форма  $R_8$  связана с радемахеровской формой  $R_4$  через кронекеровское умножение последней на матричную форму представления двойного числа (числа Лоренца) с единичными координатами ( $\otimes$  — символ кронекеровского умножения):

$$R_4 \otimes [1\ 1; 1\ 1] = R_8 \quad (1)$$

Соответственно 8-мерная алгебра с таблицей умножения на Рис. 4 справа может называться алгеброй бисплит-кватернионов Кокла, а геноматрица триплетов  $[C\ T; A\ G]^{(3)}$  в ее радемахеровском представлении  $R_8$  (Рис. 3) оказывается бисплит-кватернионом Кокла с единичными координатами. Представленные на Рис. 3 декомпозиции радемахеровских геноматриц  $R_4$  и  $R_8$  сопряжены с понятиями диадических сдвигов, диадических групп и матриц диадических сдвигов, играющих важную роль в компьютерных технологиях [7], [8, §1.2.6]. Эти декомпозиции, производимые по образцу матриц диадических сдвигов, мы называем диадо-сдвиговыми, а соответствующие матричные алгебры, обнаруживаемые при такой декомпозиции, — диадо-сдвиговыми алгебрами или ДС-алгебрами. Сплит-кватернионы и бисплит-кватернионы имеют много интересных математических свойств, ряд из которых может использоваться для математических моделей в биологии [3].

Описанная диадо-сдвиговая декомпозиция имеет прямое отношение к эволюции диалектов генетического кода (подробности см. в работе [3]). Речь идет о том, что современной науке известно целое множество природных вариантов соответствия между алфавитом 64 триплетов и алфавитом аминокислот и стоп-кодонах. Эти не сильно отличающиеся друг от друга варианты обусловлены эволюционным изменением кодового значения у некоторых из триплетов («эволюционирующих триплетов»). Существование данных вариантов (или диалектов) кода означает, что, начиная с уровня этого соответствия между этими генетическими алфавитами, наблюдается определенная эволюция генетического кода. Анализируя эволюционирующие триплеты с позиций их принадлежности или соответствия тем или иным базисным матрицам  $r_0, r_1, \dots, r_7$  генетических бисплит-кватернионов

$$\mathbf{R}_4 = \begin{array}{|c|c|c|c|} \hline + & + & + & - \\ \hline - & + & - & - \\ \hline + & - & + & + \\ \hline - & - & - & + \\ \hline \end{array} ; \quad \mathbf{R}_8 = \begin{array}{|c|c|c|c|c|c|c|c|} \hline + & + & + & + & + & + & - & - \\ \hline + & + & + & + & + & + & - & - \\ \hline - & - & + & + & - & - & - & - \\ \hline - & - & + & + & - & - & - & - \\ \hline + & + & - & - & + & + & + & + \\ \hline + & + & - & - & + & + & + & + \\ \hline - & - & - & - & - & - & + & + \\ \hline - & - & - & - & - & - & + & + \\ \hline \end{array}$$

$$\mathbf{R}_4 = c_0 + c_1 + c_2 + c_3 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{vmatrix}$$

$$\mathbf{R}_8 = r_0 + r_1 + r_2 + r_3 + r_4 + r_5 + r_6 + r_7 =$$

$$\begin{aligned}
 &= \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{vmatrix} + \\
 &+ \begin{vmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} + \\
 &+ \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}
 \end{aligned}$$

Рис. 3: Вверху: радемахеровские представления  $R_4$  и  $R_8$  геноматриц дуплетов и триплетов  $[C T; A G]^{(2)}$  и  $[C T; A G]^{(3)}$  (Рис. 2). Знаки «+» и «-» обозначают элементы «+1» и «-1». Внизу: диадо-сдвиговая декомпозиция каждого из этих радемахеровских представлений  $R_4$  и  $R_8$  в виде сумм из 4 и 8 разреженных матриц соответственно.

	<b>1</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>c<sub>3</sub></b>
<b>1</b>	<b>1</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>c<sub>3</sub></b>
<b>c<sub>1</sub></b>	<b>c<sub>1</sub></b>	<b>-1</b>	<b>c<sub>3</sub></b>	<b>-c<sub>2</sub></b>
<b>c<sub>2</sub></b>	<b>c<sub>2</sub></b>	<b>-c<sub>3</sub></b>	<b>1</b>	<b>-c<sub>1</sub></b>
<b>c<sub>3</sub></b>	<b>c<sub>3</sub></b>	<b>c<sub>2</sub></b>	<b>c<sub>1</sub></b>	<b>1</b>

;

	<b>1</b>	<b>r<sub>1</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>3</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>6</sub></b>	<b>r<sub>7</sub></b>
<b>1</b>	<b>1</b>	<b>r<sub>1</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>3</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>6</sub></b>	<b>r<sub>7</sub></b>
<b>r<sub>1</sub></b>	<b>r<sub>1</sub></b>	<b>1</b>	<b>r<sub>3</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>7</sub></b>	<b>r<sub>6</sub></b>
<b>r<sub>2</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>3</sub></b>	<b>-1</b>	<b>-r<sub>1</sub></b>	<b>r<sub>6</sub></b>	<b>r<sub>7</sub></b>	<b>-r<sub>4</sub></b>	<b>-r<sub>5</sub></b>
<b>r<sub>3</sub></b>	<b>r<sub>3</sub></b>	<b>r<sub>2</sub></b>	<b>-r<sub>1</sub></b>	<b>-1</b>	<b>r<sub>7</sub></b>	<b>r<sub>6</sub></b>	<b>-r<sub>5</sub></b>	<b>-r<sub>4</sub></b>
<b>r<sub>4</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>5</sub></b>	<b>-r<sub>6</sub></b>	<b>-r<sub>7</sub></b>	<b>1</b>	<b>r<sub>1</sub></b>	<b>-r<sub>2</sub></b>	<b>-r<sub>3</sub></b>
<b>r<sub>5</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>4</sub></b>	<b>-r<sub>7</sub></b>	<b>-r<sub>6</sub></b>	<b>r<sub>1</sub></b>	<b>1</b>	<b>-r<sub>3</sub></b>	<b>-r<sub>2</sub></b>
<b>r<sub>6</sub></b>	<b>r<sub>6</sub></b>	<b>r<sub>7</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>3</sub></b>	<b>1</b>	<b>r<sub>1</sub></b>
<b>r<sub>7</sub></b>	<b>r<sub>7</sub></b>	<b>r<sub>6</sub></b>	<b>r<sub>5</sub></b>	<b>r<sub>4</sub></b>	<b>r<sub>3</sub></b>	<b>r<sub>2</sub></b>	<b>r<sub>1</sub></b>	<b>1</b>

Рис. 4: Таблицы умножения для набора разреженных матриц из декомпозиций на Рис. 3. Слева: для радемахеровского представления  $R_4$  геноматрицы дуплетов  $[C\ T; A\ G]^{(2)}$ . Справа: для радемахеровского представления  $R_8$  геноматрицы триплетов  $[C\ T; A\ G]^{(3)}$ .

(или рассматриваемых ниже бикватернионов Гамильтона) (см. Рис. 3-6), автор обнаружил существование следующих феноменологических правил эволюции диалектов генетического кода:

**Правило №1** Абсолютным правилом для множества всех двуполых организмов (размножающихся половым путем) является то, что у него эволюционные изменения в кодовом соответствии триплетов аминокислотам и стоп-кодонам реализуются только у тех триплетов, которые сопряжены с базисными матрицами  $r_4, r_5, r_6, r_7$  генетических бисплит-кватернионов.

Это Правило №1 может быть переформулировано в виде Принципа запрета, обладающего предсказательной силой: "у двуполых организмов запрещены эволюционные изменения тех триплетов, которые сопряжены с названными базисными матрицами 8-мерных гиперкомплексных чисел".

**Правило №2** Во всех диалектах генетического кода только триплеты, сопряженные с базисными матрицами  $r_2, r_6, r_7$  генетических бисплит-кватернионов, могут быть старто-кодонами. (Отметим, что оба этих правила, включая указанный принцип запрета, с равным основанием могут формулироваться применительно к рассматриваемым ниже базисным матрицам бикватернионов Гамильтона).

### 3 Генетические матрицы Адамара

Теперь покажем, что свойства генетических алфавитов связывают генетическую систему со специальным подсемейством матриц Адамара, которые являются одним из наиболее известных инструментов в технике обработки сигналов.

4-буквенный алфавит азотистых оснований А, С, G, Т/У имеет следующие две особенности, связанные с уникальным статусом тимина Т (или урацила U):

- каждое из трех оснований А, С, G имеет функционально важную аминогруппу  $NH_2$ , а четвертое основание Т/У не имеет ее;
- буква Т — единственное основание в ДНК, которое заменяется при переходе к РНК на другое основание U.

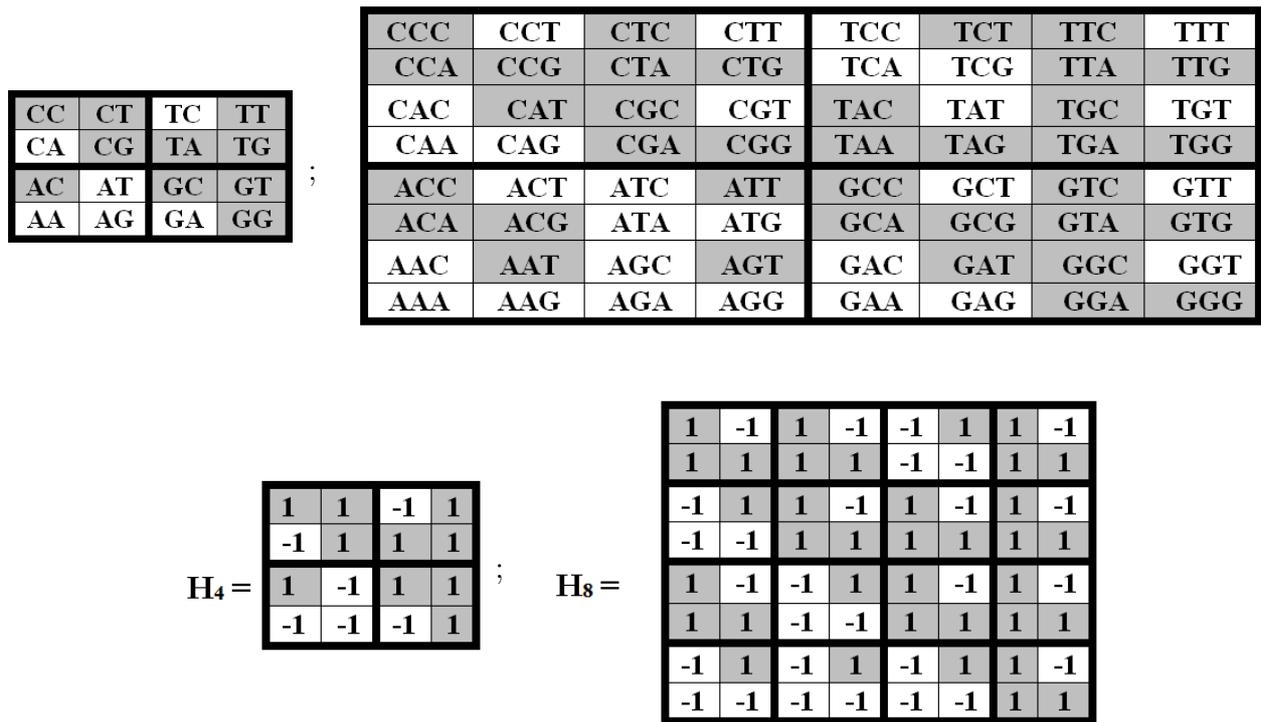


Рис. 5: Вверху: символьные мозаичные матрицы дуплетов и триплетов  $[C\ T; A\ G]^{(2)}$  и  $[C\ T; A\ G]^{(3)}$ , мозаика которых получена с помощью первого шага T/U-алгоритма из мозаики этих матриц на Рис. 2. Внизу: адамаровы представления  $H_4$  и  $H_8$  этих геноматриц дуплетов и триплетов  $[C\ T; A\ G]^{(2)}$  и  $[C\ T; A\ G]^{(3)}$ , получаемые с помощью второго шага T/U-алгоритма (пояснение в тексте).

С учетом этого уникального статуса буквы T/U нами выявлен следующий «T/U-алгоритм», который может использоваться в компьютерах организмов и который трансформирует символьные  $(4 \times 4)$ - и  $(8 \times 8)$ -матрицы дуплетов и триплетов  $[C\ T; A\ G]^{(2)}$  и  $[C\ T; A\ G]^{(3)}$  в  $(4 \times 4)$ - и  $(8 \times 8)$ -матрицы Адамара  $H_4$  и  $H_8$  (Рис. 5). Данный T/U-алгоритм состоит из двух шагов:

1. каждый из мультиплетов в черно-белых  $(4 \times 4)$ - и  $(8 \times 8)$ -геноматрицах на Рис. 2 изменяет свой цвет всякий раз, когда буква T стоит на нечетной позиции (например, стоит в триплете на первой или третьей позиции);
2. затем черные и белые дуплеты и триплеты в каждой из так полученных мозаичных матриц  $[C\ T; A\ G]^{(2)}$  и  $[C\ T; A\ G]^{(3)}$  интерпретируются как элементы «+1» и «-1» соответственно.

По определению матрицей Адамара порядка «n» называется  $(n \times n)$ -матрица  $H(n)$  с элементами «+1» и «-1», которая удовлетворяет условию  $H(n) \cdot H(n)^T = n \cdot I_n$ , где  $H(n)^T$  — транспонированная матрица, а  $I_n$  — единичная матрица. Строки или столбцы матриц Адамара образуют полную ортогональную систему функций Уолша. Приложениям матриц Адамара в технике обработки сигналов посвящены десятки тысяч публикаций: помехоустойчивое кодирование, сжатие информации, и т.д. [9]. Например, помехоустойчивые коды на основе матриц Адамара использовались на космических кораблях «Маринер» и «Вояджер», что позволило получить качественные фотографии Марса, Юпитера, Сатурна, Урана и Нептуна, несмотря на искаженность и ослабленность приходящих сигналов. Матрицы Адамара используются для создания квантовых компьютеров, которые базируются на «гейтах Адамара». В качестве унитарных операторов они применяются в квантовой механике.

Осуществим диадо-сдвиговые декомпозиции генетических матриц Адамара  $H_4$  и  $H_8$  по аналогии с декомпозициями радемахеровских матриц  $R_4$  и  $R_8$ . В результате возникают два новых набора из четырех ( $q_0, q_1, q_2, q_3$ ) и восьми разреженных матриц ( $h_0, h_1, h_2, h_3, h_4, h_5, h_6, h_7$ ), показанные на Рис. 6.

$$\begin{aligned}
 H_4 = q_0 + q_1 + q_2 + q_3 &= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{vmatrix} \\
 \\
 H_8 = h_0 + h_1 + h_2 + h_3 + h_4 + h_5 + h_6 + h_7 &= \\
 \\
 &= \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{vmatrix} + \\
 \\
 &+ \begin{vmatrix} 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} + \\
 \\
 &+ \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}
 \end{aligned}$$

Рис. 6: Диадо-сдвиговые декомпозиции геноматриц Адамара  $H_4$  и  $H_8$  из Рис. 5.

Каждый из этих двух наборов из 4 и 8 разреженных матриц неожиданно также оказывается замкнутым относительно умножения и определяет свою таблицу умножения, показанную на Рис. 7.

Левая из этих таблиц совпадает с таблицей умножения кватернионов Гамильтона, а правая — с таблицей умножения 8-мерной алгебры бикватернионов Гамильтона, которые в математике обычно определяются как кватернионы Гамильтона над полем комплексных чисел. В этой связи геноматрица Адамара  $H_4$  оказывается кватернионом Гамильтона

	<b>1</b>	<b>q<sub>1</sub></b>	<b>q<sub>2</sub></b>	<b>q<sub>3</sub></b>
<b>1</b>	<b>1</b>	<b>q<sub>1</sub></b>	<b>q<sub>2</sub></b>	<b>q<sub>3</sub></b>
<b>q<sub>1</sub></b>	<b>q<sub>1</sub></b>	<b>-1</b>	<b>q<sub>3</sub></b>	<b>-q<sub>2</sub></b>
<b>q<sub>2</sub></b>	<b>q<sub>2</sub></b>	<b>-q<sub>3</sub></b>	<b>-1</b>	<b>q<sub>1</sub></b>
<b>q<sub>3</sub></b>	<b>q<sub>3</sub></b>	<b>q<sub>2</sub></b>	<b>-q<sub>1</sub></b>	<b>-1</b>

;

	<b>1</b>	<b>h<sub>1</sub></b>	<b>h<sub>2</sub></b>	<b>h<sub>3</sub></b>	<b>h<sub>4</sub></b>	<b>h<sub>5</sub></b>	<b>h<sub>6</sub></b>	<b>h<sub>7</sub></b>
<b>1</b>	<b>1</b>	<b>h<sub>1</sub></b>	<b>h<sub>2</sub></b>	<b>h<sub>3</sub></b>	<b>h<sub>4</sub></b>	<b>h<sub>5</sub></b>	<b>h<sub>6</sub></b>	<b>h<sub>7</sub></b>
<b>h<sub>1</sub></b>	<b>h<sub>1</sub></b>	<b>-1</b>	<b>h<sub>3</sub></b>	<b>-h<sub>2</sub></b>	<b>h<sub>5</sub></b>	<b>-h<sub>4</sub></b>	<b>h<sub>7</sub></b>	<b>-h<sub>6</sub></b>
<b>h<sub>2</sub></b>	<b>h<sub>2</sub></b>	<b>h<sub>3</sub></b>	<b>-1</b>	<b>-h<sub>1</sub></b>	<b>-h<sub>6</sub></b>	<b>-h<sub>7</sub></b>	<b>h<sub>4</sub></b>	<b>h<sub>5</sub></b>
<b>h<sub>3</sub></b>	<b>h<sub>3</sub></b>	<b>-h<sub>2</sub></b>	<b>-h<sub>1</sub></b>	<b>1</b>	<b>-h<sub>7</sub></b>	<b>h<sub>6</sub></b>	<b>h<sub>5</sub></b>	<b>-h<sub>4</sub></b>
<b>h<sub>4</sub></b>	<b>h<sub>4</sub></b>	<b>h<sub>5</sub></b>	<b>h<sub>6</sub></b>	<b>h<sub>7</sub></b>	<b>-1</b>	<b>-h<sub>1</sub></b>	<b>-h<sub>2</sub></b>	<b>-h<sub>3</sub></b>
<b>h<sub>5</sub></b>	<b>h<sub>5</sub></b>	<b>-h<sub>4</sub></b>	<b>h<sub>7</sub></b>	<b>-h<sub>6</sub></b>	<b>-h<sub>1</sub></b>	<b>1</b>	<b>-h<sub>3</sub></b>	<b>h<sub>2</sub></b>
<b>h<sub>6</sub></b>	<b>h<sub>6</sub></b>	<b>h<sub>7</sub></b>	<b>-h<sub>4</sub></b>	<b>-h<sub>5</sub></b>	<b>h<sub>2</sub></b>	<b>h<sub>3</sub></b>	<b>-1</b>	<b>-h<sub>1</sub></b>
<b>h<sub>7</sub></b>	<b>h<sub>7</sub></b>	<b>-h<sub>6</sub></b>	<b>-h<sub>5</sub></b>	<b>h<sub>4</sub></b>	<b>h<sub>3</sub></b>	<b>-h<sub>2</sub></b>	<b>-h<sub>1</sub></b>	<b>1</b>

Рис. 7: Таблицы умножения для наборов из 4 и 8 разреженных матриц на Рис. 6 (матрицы  $q_0$  и  $h_0$  — единичные) для геноматриц Адамара  $H_4$  и  $H_8$ .

с единичными координатами, а геноматрица Адамара  $H_8$  — бикватернионом Гамильтона с единичными координатами. Между этими двумя геноматрицами Адамара  $H_4$  и  $H_8$  имеется алгебраическая связь через кронекеровское умножение матрицы  $H_4$  на матричное представление комплексного числа с единичными координатами:

$$H_4 \otimes [1 \quad -1; 1 \quad 1] = H_8 \tag{2}$$

Наше трехмерное физическое пространство почему-то устроено так, что соответствует по своим геометрическим особенностям кватернионам Гамильтона. А теперь оказывается, что и система генетического кодирования почему-то устроена в соответствии с кватернионами и бикватернионами Гамильтона. Кватернионы Гамильтона тесно связаны с матрицами Паули, теорией электромагнитного поля (Максвелл написал свои уравнения именно на языке кватернионов Гамильтона), специальной теорией относительности, теорией спинов, квантовомеханической теорией химической валентности и пр. Кватернионам в физике только в XX веке посвящены тысячи работ [<http://arxiv.org/abs/math-ph/0511092>]. Матричная генетика привела к открытию важного моста между физикой, биологией и информатикой для их взаимного обогащения.

В наших исследованиях получен новый пример непостижимой эффективности математики: абстрактные математические структуры, выведенные математиками на кончике пера 160 лет назад, оказываются воплощенными в информационной основе живой материи — системе генетических алфавитов. И то, что математики открывают путем мучительных раздумий (подобно Гамильтону, потратившему 10 лет непрерывных раздумий для открытия его кватернионов), оказывается уже представленным в системе генетического кодирования. Наша генетическая система соответствует нашему физическому пространству, поскольку — как и оно — соответствует кватернионам Гамильтона. Это, в частности, затрагивает давно дискутируемый в науке вопрос о врожденности представлений животных организмов о пространстве [10, с. 77-80].

Кватернионы Гамильтона и сплит-кватернионы Кокла объединены тем, что весь набор алгебр Клиффорда для двумерного случая  $Cl_{1,1}(\mathbb{R})$ ,  $Cl_{2,0}(\mathbb{R})$ ,  $Cl_{0,2}(\mathbb{R})$  изоморфен именно этим двум 4-мерным алгебрам гиперкомплексных чисел ([http://www.mi.ras.ru/noc/11\\_12/cllifalg04.12.11.pdf](http://www.mi.ras.ru/noc/11_12/cllifalg04.12.11.pdf)). Это интересно для международного сообщества «клиффордистов», стремящихся трактовать алгебры Клиффорда как универсальный язык геометрии и физики.

Кватернионы Гамильтона известны как инструмент математики и физики. Нами обращается внимание на то, что кватернион Гамильтона с единичными координатами через

<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
<b>-b</b>	<b>a</b>	<b>-d</b>	<b>c</b>
<b>-c</b>	<b>d</b>	<b>a</b>	<b>-b</b>
<b>-d</b>	<b>-c</b>	<b>b</b>	<b>a</b>

;

<b>a</b>	<b>b</b>	<b>-c</b>	<b>d</b>
<b>-b</b>	<b>a</b>	<b>d</b>	<b>c</b>
<b>c</b>	<b>-d</b>	<b>a</b>	<b>b</b>
<b>-d</b>	<b>-c</b>	<b>-b</b>	<b>a</b>

;

<b>a</b>	<b>-b</b>	<b>c</b>	<b>d</b>
<b>b</b>	<b>a</b>	<b>-d</b>	<b>c</b>
<b>-c</b>	<b>d</b>	<b>a</b>	<b>b</b>
<b>-d</b>	<b>-c</b>	<b>-b</b>	<b>a</b>

Рис. 8: Примеры матричных представлений кватернионов Гамильтона  $a * 1 + b * i + c * j + d * k$ . При единичных значениях координат ( $a = b = c = d = 1$ ) все эти матрицы являются матрицами Адамара. Расположение координат  $a, b, c, d$  в каждой матрице соответствует принципу строения матриц диадических сдвигов, описанных, например, в [7].

его матричные представления ( $a$  существует не одно такое представление) оказывается глубоко связанным с информатикой в силу двух причин:

1. его  $(4 * 4)$ -матричные представления являются матрицами Адамара;
2. размещение кватернионных координат четырех видов в любых его  $(4 * 4)$ -матричных представлениях всегда соответствует упомянутому принципу диадических сдвигов, основанному на одной из основных операций компьютерных технологий — сложении бинарных чисел по модулю 2.

Рис. 8 демонстрирует эти свойства на конкретных примерах  $(4 * 4)$ -матричных представлений кватернионов Гамильтона.

Почему кватернионы и бикватернионы Гамильтона оказываются связанными с системой генетического кодирования? Вероятно, глубокий ответ на этот вопрос будет получен при учете не только геометрических свойств кватернионов и бикватернионов Гамильтона (например, их соответствия поворотам нашего физического 3D-пространства), но и их информационной сущности и пригодности для эффективной обработки информации в качестве особого вида матриц Адамара со специальными системами функций Уолша в них и пр.

Укажем возможные причины того, почему молекулярно-генетическая система алфавитов оказывается связанной с ассоциативными алгебрами в отличие от систем теоретической физики неживой материи, где большое внимание уделяется неассоциативным алгебрам октонионов Кэли и сплит-октонионов. Предполагаемая причина заключается в том, что ассоциативность является важным условием в информатике помехоустойчивого кодирования, где широко используются многие виды кодов (например, циклические коды), которые базируются на концепции алгебраических групп и полей Галуа. При этом сами определения понятий группы и полей Галуа содержат в себе условия ассоциативности их элементов. Тем самым, в области помехоустойчивого кодирования ассоциативные матричные алгебры гиперкомплексных чисел обладают фундаментальным преимуществом по сравнению с неассоциативными алгебрами октонионов Кэли и сплит-октонионов.

Названные кватернион и бикватернион Гамильтона с единичными координатами, будучи нормированы на единичный детерминант, образуют при возведении в степени циклические группы с периодами 6 и 24 соответственно:

$$(2^{-1} * H_4)^n = (2^{-1} * H_4)^{n+6}; \quad (2^{-1.5} * H_8)^n = (2^{-1.5} * H_8)^{n+24} \quad (3)$$

Свойства этих циклических групп таковы, что позволяют использовать их для моделирования генетически наследуемых свойств живых организмов. Например, покажем совпадение свойств генетически наследуемого цветовосприятия с 6-членной циклической группой

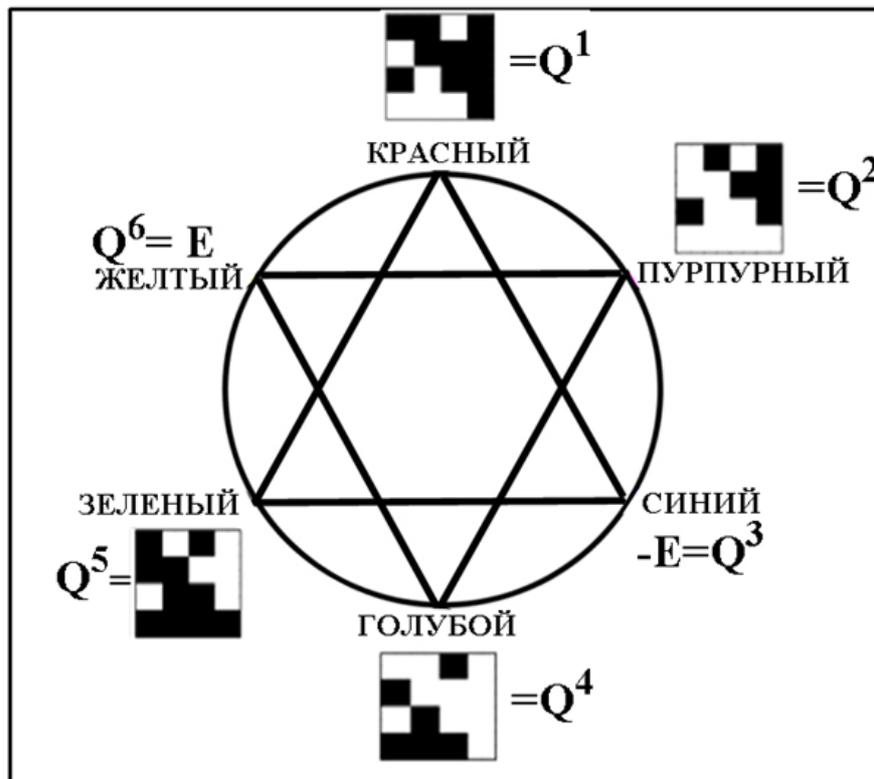


Рис. 9: Цветовой круг Ньютона из психофизики цветовосприятия и соответствие ему членов циклической группы нормированного кватерниона Гамильтона с единичными координатами  $Q^n = (0.5 * H_4)^n$ , период которой равен 6.

кватерниона Гамильтона  $(2^{-1} * H_4)^n$ . Изобразим все члены этой циклической группы на круговом циферблате (черные ячейки матриц содержат числа «+0, 5», белые — «-0, 5»; E — единичная матрица) (Рис. 9). Эта 6-членная группа кватернионов Гамильтона имеет следующие свойства:

1. Противолежащие на круге кватернионы противоположны по знаку и при сложении дают ноль, а потому являются взаимно дополнительными;
2. Каждый кватернион на круге представляет собой сумму двух соседних кватернионов;
3. Сумма кватернионов в вершинах каждого из двух треугольников «звезды Давида» равна нулю.

Эти свойства циклической группы реперного кватерниона Гамильтона тождественны феноменологическим свойствам наследуемого цветовосприятия. Последнее основано на трех базовых цветах (красный, синий, зеленый) и трех дополнительных к ним (голубой, желтый, пурпурный), лежащих в вершинах двух треугольников «звезды Давида» на цветовом круге Ньютона:

1. противолежащие на круге цвета взаимно дополнительные и при наложении гасят друг друга;
2. каждый цвет на круге представляет собой сумму двух соседних цветов;
3. три базовые цвета, как и три дополнительные цвета, при наложении гасят друг друга.

Задачу о смешении цветов теперь формально можно решать на языке циклической группы кватерниона Гамильтона, если закрепить за каждым цветом член этой группы в соответствии с Рис. 9. Например, какой цвет получится, если смешать 3 доли красного

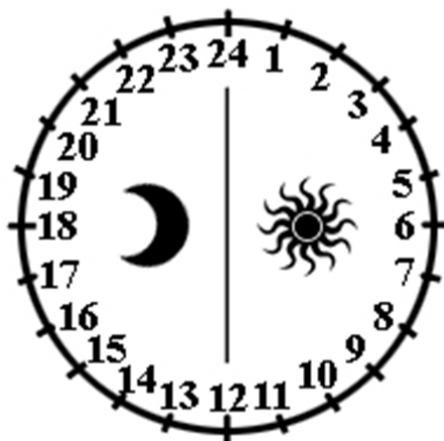


Рис. 10: 24-часовой циферблат суток с их делением на дневную и ночную 12-часовые половины.

цвета, 2 доли желтого и 5 долей синего? Ответ получается сложением соответствующих кватернионов:  $3Q^1 + 2Q^6 + 5Q^3 = 3Q^2$ , т.е. 3 доли пурпурного (здесь использовано обозначение  $Q = 0.5 * H_4$ ).

В психофизике известно, что цвет — это не физическое свойство объекта, а наследуемая реакция человека на поступающее от объекта световое воздействие с учетом общего светового окружения. Почему же наше цветовое восприятие устроено именно в соответствии с цветовым кругом и циклической группой кватернионов Гамильтона? Возможный ответ: потому что организм обрабатывает световую информацию на основе наследуемых единых алгоритмов обработки генетической и физической информации с участием генетически наследуемой циклической группы кватернионов Гамильтона, с которой связана сама система генетических алфавитов.

Свойства указанной выше (3) циклической группы бикватерниона Гамильтона с периодом 24 позволяют использовать ее для математического моделирования в области проблематики 24-часовых биоритмов суток, тем более, что бикватернионы в физике используются для моделирования пространственно-временных отношений. Как известно, наш организм представляет собой огромный хор генетически наследуемых ритмических процессов. Даже белки циклически распадаются на аминокислоты и собираются вновь. Фазы активности и пассивности физиологических систем согласованы между собой и по представлениям древней и современной медицины обуславливают деление человеком суток на 24 часа. Биоритмическая организация наследуемых процессов распределена по всему телу организмов самых различных видов, в том числе не обладающих нервной системой. Биоритмология важна для медицины, эргономики и пр. (см., например, [11]).

Рис. 10 показывает 24-часовой циферблат, в котором числа соответствуют показателям степени этого нормированного бикватерниона ( $B = 2^{-1.5} * H_8$ ) с единичными координатами. Бикватернионы в степенях, противоположащих на циферблате, отличаются только противоположным знаком ( $B^n = -B^{n+12}$ ) по аналогии с противоположностью дневных и ночных часов (например, в обычной жизни вместо «14 часов» зачастую говорят «2 часа дня», называя противоположащее на этом циферблате число). Эта циклическая группа бикватернионов имеет также другие интересные свойства, например:  $B^n = B^{n-4} + B^{n+4}$ ;  $B^n = -(B^{n-8} + B^{n+8})$ ;  $B^n = 2^{-0.5} * (B^{n-3} + B^{n+3})$ .

Теория обработки сигналов уделяет особое внимание перестановкам информационных элементов. В матричной генетике нами изучены все 6 возможных вариантов одновременной перестановки позиций в триплетах: 1 – 2 – 3, 2 – 3 – 1, 3 – 1 – 2, 3 – 2 – 1, 2 – 1 – 3, 1 – 3 – 2. Такие перестановки трансформируют большинство триплетов в ячейках исходной

матрицы триплетов  $[C T; A G]^{(3)}$ . Например, в случае циклической трансформации порядка позиций  $1 - 2 - 3$  в порядок  $2 - 3 - 1$ , триплет  $CAG$  со слабым корнем  $CA$  трансформируется в триплет  $AGC$  с сильным корнем  $AG$ . В результате возникают новые мозаичные матрицы триплетов. При таких перестановках исходная радемахеровская геноматрица  $R$  трансформируется в 5 новых радемахеровских матриц, а исходная геноматрица Адамара  $H$  трансформируется в 5 новых геноматриц Адамара [3].

Соответствующая диадо-сдвиговая декомпозиция каждой из пяти новых радемахеровских матриц приводит к новому набору из 8 разреженных матриц, который также оказывается замкнутым относительно умножения и соответствует той же самой таблице умножения бисплит-кватернионов Кокла (Рис. 4). Это же справедливо для диадо-сдвиговой декомпозиции каждой из пяти новых геноматриц Адамара, которая приводит к новому набору из 8 разреженных матриц, замкнутому относительно умножения и соответствующему той же самой таблице умножения бикватернионов Гамильтона на Рис. 7. Аналогичная алгебраическая инвариантность имеет место для  $(4 * 4)$ -геноматриц дуплетов  $[C T; A G]^{(2)}$  при перестановке позиций в дуплетах: перейдя от порядка позиций  $1 - 2$  к порядку  $2 - 1$ , мы получаем новую по мозаике  $(4 * 4)$ -геноматрицу, радемахерово и адамарово представления которой являются новыми матричными представлениями тех же сплит-кватерниона Кокла и кватерниона Гамильтона с единичными координатами [3].

Вопрос об инвариантности матричных алгебр при разных перестановках элементов в геноматрицах интересен еще и в связи с метаморфозами организмов. Например, в метаморфозе бабочки, включающим четыре стадии (яйцо – гусеница – куколка – бабочка), куколка ничем не питается и имеет фиксированный атомарный состав, но за счет генетически детерминированной перестановки элементов куколка чудесным образом превращается в новый организм — бабочку.

Подчеркнем, что речь в статье идет только о подмножествах кватернионов и сплит-кватернионов с единичными координатами, а также об их целых степенях, но не об их полных алгебрах над всем полем вещественных или комплексных чисел, к которым привыкли физики. Автор полагает, что в полном виде эти алгебры над полем вещественных или комплексных чисел в биологии не реализуются. По ряду данных можно думать о том, что для биоматематики данные алгебры и их матричные расширения пригодны при их определении над полем алгебраических чисел.

На основе выявленных связей системы генетических алфавитов с описанными в статье гиперкомплексными числами автором устанавливаются новые количественные и симметрические правила организации длинных ДНК секвенций нуклеотидов (в дополнение к известному «второму правилу Чаргафа»). Одновременно полученные результаты выводят на новые средства анализа генетически наследуемых физиологических процессов. Речь идет об использовании названных геноматриц Адамара в связи с плодотворной секвентной теорией Х. Хармута обработки сигналов [8, 12-14]. Она эффективно используется в телекоммуникациях, радарной технике и пр. Генетические матрицы Адамара определяют специальные варианты секвентного спектрального анализа, пространственной и временной фильтрации физиологических процессов и их корреляционного анализа. Данные развиваемой «матричной генетики» позволяют полагать, что этот «генетический» секвентный анализ может послужить ключом к важным аспектам биоинформатики.

Широкое использование матриц Адамара и их систем функций Уолша в цифровой информатике определяется их особыми свойствами. Во-первых, ступенчатые функции Уолша реализуются в электронных устройствах много проще, чем тригонометрические и многие другие функции. Во-вторых, использование функций Уолша позволяет осуществлять обработку цифровых сигналов посредством одних только операций сложения и вычитания без использования умножения и деления (см., например, [7, 15]). Поскольку в этом случае операций деления не требуется, то в цифровой информатике могут использоваться

алгебры без деления, к числу которых относятся расширения алгебр кватернионов и сплит-кватернионов на случаи более высоких  $2^n$ -мерных алгебр. Это дополнительно отличает компьютерную информатику от теоретической физики, где операции умножения и деления существенны и где в этой связи внимание физиков обращено на алгебры с делением, прежде всего, на алгебры октонионов Кэли. Для рассматриваемой нами области молекулярной генетики существенно, что в ней операции сложения и вычитания могут быть организованы посредством простого соединения или разъединения молекулярных элементов, что, очевидно, много проще организации операций умножения и деления на молекулах. Следует отметить существование некоторых работ по приложению функций Уолша с их полезными свойствами для спектрального анализа генетических секвенций и создания новых генетических алгоритмов [16-22].

Известно, что в области многоканальной цифровой связи многомерные числовые системы используются для обеспечения функциональных преимуществ. Например, 2-мерные комплексные числа используются для цифровой обработки 2-мерных сигналов [23, гл. 8]; в этом случае вещественная часть комплексных чисел соответствует сигналу первого канала, а мнимая часть комплексных чисел — сигналу второго канала. Но биологические организмы представляют собой многоканальные информационные системы. Например, сетчатка глаза обладает врожденным набором рецепторов, который осуществляет многоканальную передачу в нервную систему информации о проецируемых на сетчатку образцов. Но для эффективной работы информационных систем с большим числом параллельных каналов передачи потоков независимых исходных сигналов требуется кооперативный принцип учета этих потоков. Это принцип может быть реализован в форме алгебраической системы гиперкомплексных чисел. Мы полагаем, что описанные генетические гиперкомплексные числа используются в наследуемых многоканальных биоинформационных системах для подобных целей.

Молекулярная генетика располагает примерами особой феноменологической роли числа 8. Например, в клетках эукариотов нити ДНК навиты вокруг нуклеосом, каждая из которых представляет собой октамерный стержень, состоящий из гистоновых белков четырех типов: H2A, H2B, H3 and H4. Точнее, одиночная нуклеосома является ансамблем из 8 гистонов, в котором присутствует пара гистонов каждого из этих четырех видов (между этими гистонами имеются отношения комплементарности, напоминающие отношения комплементарности между парами азотистых оснований в ДНК). Молекула ДНК навита на этот октамерный стержень. Октамерная структура нуклеосом играет центральную роль в упаковке ДНК на всех уровнях. В молекулярной генетике существует широко известная концепция гистонового кода [24]. Возможно, что математические формализмы описанных генетических гиперкомплексных чисел окажутся полезными для моделирования феноменологических фактов этой фундаментальной концепции и для ее развития.

Можно добавить, что в области молекулярной генетики для обеспечения тех или иных физиологических процессов должны согласованно работать целые семейства белков (отдельный, изолированный белок не является живой системой). При этом данные семейства должны содержать определенные количества белков разных видов. Генетическая система успешно решает эту числовую задачу генетической детерминации количеств белков каждого типа внутри отдельных семейств для разных биологических процессов. Описанные в статье генетические гиперкомплексные числа могут быть дополнительно использованы для математического моделирования подобных генетических феноменов. Обобщенно говоря, можно думать, что в области молекулярной генетики и наследуемых физиологических структур эти генетические гиперкомплексные системы могут служить естественным генетическим базисом для обеспечения числовых закономерностей в наследуемых множествах информационных элементов.

## 4 Решетки Пеннета, кронекеровские произведения и ДС-декомпозиции

Структурные особенности различных уровней генетической системы должны быть взаимно скоординированы для обеспечения единства организма и наследуемой передачи физиологических подсистем по цепи поколений. В этой связи методы матричной генетики могут быть полезны при их приложении к анализу и моделированию различных уровней генетической организации в живом.

До сих пор мы говорили об алфавитах генетических мультиплетов и их субалфавитах. Этот параграф посвящен новому классу объектов матричной генетики, относящихся к более высокому биологическому уровню, где новые термины фигурируют в качестве центральных: «аллели», «гаметы», «генотип», «зигота», и пр. Мы будем анализировать классические для генетики квадратные решетки Пеннета. Они представляют алфавиты генотипов или, точнее, алфавиты гибридов в возможных комбинациях мужских и женских гамет (половых клеток) при скрещивании организмов в соответствии с законами Менделя. Мы покажем возможность интерпретации решеток Пеннета для полигибридного скрещивания не как таблиц, а как квадратных матриц (будем называть их «матрицами Пеннета»), получаемых в результате кронекеровского умножения  $(2 * 2)$ –матриц Пеннета для моногибридного скрещивания. Основываясь на кронекеровском умножении матриц, использованном нами на микроуровне молекулярно-генетических алфавитов, этот матричный подход оказывается пригодным для феноменологии биообъектов значительно более высокого уровня: он дает простой алгебраический способ конструирования решеток Пеннета для мультигибридного скрещивания организмов. Дополнительно мы покажем, что диадо-сдвиговые декомпозиции (ДС-декомпозиции) этих «матриц Пеннета» ведут в некоторых важных случаях к методу классификации различных подмножеств комбинации аллелей от женских и мужских гамет. Некоторые из этих результатов матричной генетики на уровне генотипов формально идентичны результатам, описанным выше для более низкого уровня генетических мультиплетов. Эти результаты демонстрируют алгебраическую общность генетической организации на различных уровнях.

Решетки Пеннета являются широко известным инструментом в генетике. Они были введены английским генетиком Пеннетом (R. C. Punnett) в 1905 году для предсказания генетического наследования признаков в потомках. Этот метод Пеннета излагается во множестве учебников по генетике и на сайтах Интернета.

Наследование заключается в передаче признаков от родителей к потомкам. Передаваемые признаки контролируются генами. Различные формы или состояния гена для того или иного определенного признака называются аллелями. Для каждого признака обычно имеется два аллеля (или два состояния гена). Аллели могут быть доминантными или рецессивными. Каждая клетка в организме содержит два аллеля для каждого из признаков. Один аллель наследуется от материнского организма, а второй – от отцовского. Решетка Пеннета дает простой метод предсказания комбинаций этих аллелей при скрещивании организмов. Решетка Пеннета представляет все возможные комбинации материнских аллелей с отцовскими аллелями при полигибридном скрещивании. В решетках Пеннета доминантные и рецессивные аллели принято обозначать буквами: заглавные буквы обозначают доминантные аллели (мы будем использовать для доминантных аллелей буквы H, B, C, ...), а строчные буквы обозначают рецессивные аллели (соответственно h, b, c, ...). Организм называется гомозиготным по данному признаку (или признакам), если он имеет одинаковые аллели для этого признака, например, HH или hh. Организм называется гетерозиготным, если он имеет не одинаковые аллели для данного признака, например, Hh. Для каждого наследуемого признака существует три возможных комбинации аллелей: гомозиготная доминантная (HH), гетерозиготная (Hh) и гомозиготная рецессивная (hh). При

		материнские гаметы							
		HbC	Hbc	HbC	Hbc	hbC	hbc	hbC	hbc
отцовские гаметы	HBC	<u>HHBbCC</u>	<u>HNBbCe</u>	<u>HNBbCC</u>	<u>HNBbCe</u>	<u>HhBbCC</u>	<u>HhBbCe</u>	<u>HhBbCC</u>	<u>HhBbCe</u>
	HBe	<u>HNBbCe</u>	<u>HNBbcc</u>	<u>HNBbCe</u>	<u>HNBbcc</u>	<u>HhBbCe</u>	<u>HhBbcc</u>	<u>HhBbCe</u>	<u>HhBbcc</u>
	HbC	<u>HHbbCC</u>	<u>HHbbCe</u>	<u>HHbbCC</u>	<u>HHbbCe</u>	<u>HhbbCC</u>	<u>HhbbCe</u>	<u>HhbbCC</u>	<u>HhbbCe</u>
	Hbe	<u>HHbbCe</u>	<u>HHbbcc</u>	<u>HHbbCe</u>	<u>HHbbcc</u>	<u>HhbbCe</u>	<u>Hhbbcc</u>	<u>HhbbCe</u>	<u>Hhbbcc</u>
	hBC	<u>HhBbCC</u>	<u>HhBbCe</u>	<u>HhBbCC</u>	<u>HhBbCe</u>	<u>hhBbCC</u>	<u>hhBbCe</u>	<u>hhBbCC</u>	<u>hhBbCe</u>
	hBe	<u>HhBbCe</u>	<u>HhBbcc</u>	<u>HhBbCe</u>	<u>HhBbcc</u>	<u>hhBbCe</u>	<u>hhBbcc</u>	<u>hhBbCe</u>	<u>hhBbcc</u>
	hbC	<u>HhbbCC</u>	<u>HhbbCe</u>	<u>HhbbCC</u>	<u>HhbbCe</u>	<u>hhbbCC</u>	<u>hhbbCe</u>	<u>hhbbCC</u>	<u>hhbbCe</u>
	hbe	<u>HhbbCe</u>	<u>Hhbbcc</u>	<u>HhbbCe</u>	<u>Hhbbcc</u>	<u>hhbbCe</u>	<u>hhbbcc</u>	<u>hhbbCe</u>	<u>hhbbcc</u>

Рис. 11: Пример решетки Пеннета для тригибридного скрещивания при материнском генотипе HhbbCc и отцовском генотипе HhBbCc (из <http://www.changbioscience.com/genetics/punnett.html>).

классическом конструировании квадратной решетки Пеннета на одной ее стороне (обычно верхней) выписывают аллели материнских гамет, а на другой стороне (левой) выписывают аллели отцовских гамет; варианты генотипов возможного потомства записываются в ячейки решетки посредством взятия одного аллеля сверху (он является символом столбца этой ячейки) и одного аллеля слева (он является символом строки). По традиции заглавные буквы доминантного аллеля ставятся перед строчными буквами рецессивного аллеля (например, Hh, но не hH).

Если при скрещивании рассматривается только один наследуемый признак, то оно называется моногибридным скрещиванием. Если же рассматриваются два или три наследуемых признака, то скрещивание называется соответственно дигибридным или тригибридным. Рисунок 11 показывает пример решетки Пеннета для одного из случаев тригибридного скрещивания родителей с материнским генотипом HhbbCc и отцовским генотипом HhBbCc (этот частный пример взят с сайта о решетках Пеннета <http://www.changbioscience.com/genetics/punnett.html>). В данном примере каждый набор родительских гамет включает 8 гамет, при этом набор материнских гамет отличается от набора отцовских гамет. Следует отметить, что конструирование решетки Пеннета для мультигибридного скрещивания организмов названным традиционным путем является достаточно трудоемкой задачей, быстро усложняющейся при возрастании числа учитываемых признаков в мультигибридном скрещивании.

Нами обращается внимание на то, что эта  $(8 \times 8)$ –матрица Пеннета (заключена в жирную рамку на рисунке 11) идентична результату кронекеровского произведения трех  $(2 \times 2)$ –матриц Пеннета для моногибридного скрещивания по каждому из этих трех признаков в отдельности. Рисунок 12 показывает три решетки Пеннета для моногибридного скрещивания в рассматриваемом случае материнского генотипа HhbbCc и отцовского генотипа HhBbCc.

Случаи мультигибридного скрещивания при одинаковых материнском и отцовском генотипах описываются решетками Пеннета, которые в некоторых отношениях связаны с диадо-сдвиговыми декомпозициями [3]. Например, в случае тригибридного скрещивания материнского генотипа HhBbCc и отцовского генотипа HhBbCc  $(8 \times 8)$ –матрица Пеннета с помощью ДС-декомпозиции представляется суммой восьми разреженных  $(8 \times 8)$ –матриц  $P_0, P_1, P_2, \dots, P_7$ , каждая из которых имеет по 8 не нулевых ячеек. Тогда первая из этих

материнские гаметы		
Н	h	
Н	НН	<u>Нh</u>
h	<u>Нh</u>	<u>hh</u>

материнские гаметы		
b	b	
B	Bb	Bb
b	bb	bb

материнские гаметы		
C	c	
C	CC	Cc
c	Cc	cc

Рис. 12: Три решетки Пеннета моногибридного скрещивания для построения решетки Пеннета тригибридного скрещивания (рисунок 11) с помощью их кронекеровского произведения.

восьми матриц  $p_0$  имеет 8 не нулевых ячеек, которые содержат все 8 гомозиготных комбинаций аллелей ННВСС, ННВсс, ННЬЬСС, ННЬЬсс, hhВСС, hhВсс, hhЬЬСС, hhЬЬсс. Вторая из этих матриц  $p_1$  содержит только все комбинации аллелей, в которых только третий наследуемый признак гетерозиготен: ННВССс, ННВССс, ННЬЬСс, ННЬЬСс, hhВССс, hhВССс, hhЬЬСс, hhЬЬСс, и т.д. Другими словами, посредством ДС-декомпозиций мы получаем метод классификации различных подмножеств гомозиготных и гетерозиготных потомков при таком мультигибридном скрещивании.

Описанные аналогии между матрицами генетических алфавитов и матрицами Пеннета для наследуемых признаков позволяют автору выдвигать рабочую гипотезу «алфавитно-молекулярного менделизма». Согласно этой гипотезе надхромосомные феномены наследования признаков у целостных живых организмов, которые были выявлены Менделем и подчиняются сформулированным им законам, возникли не на пустом месте, но являются продолжением молекулярных феноменов, которые определены более глубокими законами молекулярно-алфавитного уровня. Эти более глубокие законы имеют существенные аналогии с законами Менделя. Множество наследуемых признаков в биологических организмах может рассматриваться как особая алфавитная система высокого уровня, а сам организм — как некоторый текст, набранный в этой алфавитной системе. С этой точки зрения, биологическая эволюция предстает в определенной степени как эволюция многоуровневой системы взаимно связанных биологических алфавитов, начиная, по крайней мере, с молекулярно-генетического уровня. Автор предполагает существование универсальных био-алгоритмов эволюционного продуцирования взаимно связанных алфавитов все более и более высокого уровня биологической организации. Кронекеровские произведения генетических матриц, матричные формы представления генетических гиперкомплексных систем, алгоритмы диадических сдвигов и другие алгоритмы, выявляемые в наших работах по матричной генетике, могут использоваться для моделирования этих эволюционных процессов усложнения многоуровневой системы биологических алфавитов. При этом вновь образуемые алфавиты также имеют сопряжение с диадическими сдвигами, что условно может быть выражено кратким тезисом, выдвигаемым автором: «живые организмы являются диадо-сдвиговыми сущностями».

Дополнительными свидетельствами в пользу этого тезиса являются следующие факты. Матрицы диадических сдвигов крестообразны по своему строению и фракталоподобны: квадранты вдоль каждой из двух матричных диагоналей тождественны по своему числовому содержанию (то же самое относится к строению отдельных квадрантов, субквадрантов и т.д., что говорит о фракталоподобной структуре этих матриц). Но генетически наследуемое устройство наших информационных систем также крестообразно. Так, левое полушарие головного мозга обслуживает правую половину тела, а правое полушарие — левую половину. Крестообразны также зрительные нервные пути из двух глаз: эти пути передают информацию о правой половине поля зрения в левое полушарие мозга, а информацию о левой половине поля зрения — в правое полушарие. Это же верно для слуховой системы.

Настоящая статья свидетельствует, что гиперкомплексные числа оказываются важными не только для геометрии и физики, но и для биологии. Дополнительно отметим, что система генетических алфавитов сопряжена с символьной системой алфавитов древнекитайской «Книги Перемен», написанной несколько тысяч лет назад. Например, в этой древнекитайской системе базовое значение имеют четыре диграммы, именуемые Старый Ян, Старая Инь, Молодой Ян и Молодая Инь. По аналогии с генетическим алфавитом из этих диграмм можно сконструировать квадратную матрицу, третья кронекеровская степень которой порождает знаменитую таблицу 64 гексаграмм порядка Фу-си их этой древнекитайской книги [2-4]. Существует много других аналогий между генетической системой и системой «Книги Перемен». Древние китайцы утверждали, что их система является всеобщим природным архетипом и универсальной классификационной схемой. Они ничего не знали о генетическом коде, но генетический код оказывается сконструированным во многом в соответствии с этой древней системой. При этом благодаря разработкам в области матричной генетики сама система «Книги перемен» предстает как алгебраически содержательная многоуровневая система взаимно связанных алфавитов.

## Заключение

Многомерные векторные пространства и соответствующие им гиперкомплексные числовые системы интересны не только для теоретической физики неживой материи, но также для живой материи, глубоко связанной с информатикой, которая давно использует гиперкомплексные числа в цифровой технике.

Как подчеркивается в статье «Проблема происхождения жизни и мышления с точки зрения современной физики» Д.С. Чернавского («Успехи физических наук», 2000 г., № 2): *«В физике неживой природы эти понятия /информации и ценной информации/ практически не используются, поскольку они там не нужны. В биологии, напротив, понятие «информация» и особенно «ценная информация» являются основными, и без них ни понять, ни описать явления в живой природе невозможно. В этом собственно и заключается специфика «живого»».*

Именно на пути разработки информационных подходов к живой материи с использованием алгебр гиперкомплексных систем можно ожидать прорыва в ее понимании, включая возможное появление новых законов, о которых писал Э. Шредингер [25, заключительная глава]: *«Для организма следует ожидать новых законов. . . все известное нам о структуре живой материи заставляет ожидать, что деятельность живого организма нельзя свести к появлению обычных законов физики. И не потому, что имеется какая-нибудь «новая сила» или что-либо еще, управляющее поведением отдельных атомов внутри живого организма, а потому, что его структура отличается от всего изученного нами до сих пор в физической лаборатории. . . Развертывание событий в жизненном цикле организма обнаруживает удивительную регулярность и упорядоченность, не имеющих себе равных среди всего, с чем мы встречаемся в неодушевленных предметах. . . Короче говоря, мы видим, что существующая /в организме/ упорядоченность проявляет способность поддерживать саму себя и производить упорядоченные явления. . . Мы вправе предполагать, что живая материя подчиняется новому типу физического закона».*

Полученные в матричной генетике результаты заставляют автора не соглашаться с известным утверждением: *«Энциклопедия Жизни написана 4 буквами».* Это утверждение скрывает ключевую роль многоуровневой системы сопряженных генетических алфавитов. Много более верным выглядит следующее утверждение: *«Энциклопедия Жизни написана сплетением многих взаимосвязанных алфавитов, имеющим алгебраическую основу».*

Биоинформатика в широком смысле призвана решить множество научных проблем о

наследуемых свойствах организмов, включая следующие:

- помехоустойчивости генетического кодирования;
- управления и синхронизации огромного множества наследуемых циклических процессов, зачастую дихотомического типа;
- сжатия наследуемой биологической информации;
- системной организации согласованной сборки и разборки белков с их особенностями пространственного строения;
- наследования морфогенетических конструкций и процессов, и пр.

Можно полагать, что изложенные биоматематические результаты будут полезны при решении многих из этих сложных вопросов.

В начале статьи было отмечено, что живой организм является алгоритмической машиной многоканального помехоустойчивого кодирования, которая реализует законы Менделя независимого наследования признаков и пр. Представляется, что нами открыты некоторые важные алгебраические основы этой алгоритмической генетической машины, которые свидетельствуют о глубокой алгебраической природе генетической системы. Данные результаты ведут к развитию алгебраической биологии. Система генетических алфавитов своими алгебраическими особенностями подсказывает новый — алгебраический — путь к познанию живой материи.

## Благодарности

Автор выражает благодарность дирекции Института машиноведения РАН (директор — академик Р.Ф. Ганиев) за поддержку данного направления исследований, а также молодым сотрудникам лаборатории и студентам МФТИ, участвующим в его разработке: И.В. Степаняну, В.И. Свиринову, Д.Ш. Катанову, А.В. Буракову, В.И. Демо.

## Литература

- [1] Stewart I. Life's other secret: The new mathematics of the living world, New-York, Penguin, 1999.
- [2] Петухов С.В. Матричная генетика, алгебры генетического кода, помехоустойчивость. М., РХД, 2008, 316 с.
- [3] Petoukhov S.V. The genetic code, 8-dimensional hypercomplex numbers and dyadic shifts. pp. 1-31, 2011. from <http://arxiv.org/abs/1102.3596>.
- [4] Petoukhov S.V., He M. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. Hershey, USA, IGI Global. 271 p, 2010
- [5] Karlin S., Ost F., Blaisdell B.E. Patterns in DNA and amino acid sequences and their statistical significance, in: Mathematical methods for DNA sequences, M.S. Waterman, Ed., Florida: CRC Press, 1999.
- [6] Конопельченко Б.Г., Румер Ю.Б. Классификация кодонов в генетическом коде. ДАН СССР, т. 223, № 2, с. 471-474, 1975.
- [7] Ахмед Н., Рао К.Р. Ортогональные преобразования при обработке цифровых сигналов. М., 1980.
- [8] Хармут Х. Теория секвентного анализа. М., Мир, 1980
- [9] Seberry J., Wysocki B.J., Wysocki T.A. On some applications of Hadamard matrices // *Metrica*, vol. 62, pp. 221-239, 2005.

- [10] Петухов С.В. Биомеханика, бионика и симметрия. М., Наука, 1981.
- [11] Райт К. Время нашей жизни. // *В мире науки*, № 1, с. 45–51, 2003.
- [12] Хармут Х. Передача информации ортогональными функциями. М., Связь, 1975.
- [13] Хармут Х. Несинусоидальные волны в радиолокации и радиосвязи. М., Радио и связь, 1985.
- [14] Хармут Х. Применение методов теории информации в физике. М., Мир, 1989.
- [15] Трахтман А.М. Введение в обобщенную спектральную теорию сигналов. М., Советское радио, 1972.
- [16] Forrest S., Mitchell M. The performance of genetic algorithms on Walsh polynomials: Some anomalous results and their explanation. – In R.K.Belew and L.B.Booker, editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pp. 182-189. Morgan Kaufmann, San Mateo, CA, 1991.
- [17] Geadah Y.A., Corinthios M.J. Natural, dyadic and sequency order algorithms and processors for the Walsh-Hadamard transform. // *IEEE Trans. Comput.* Vol. C-26, pp. 435-442, 1977.
- [18] Goldberg D.E. Genetic algorithms and Walsh functions. // *Complex systems*. 3(2), pp. 129-171, 1989.
- [19] Lee M.H., Kaveh M. Fast Hadamard transform based on a simple matrix factorization. // *IEEE Trans. Acoust., Speech, Signal Processing*. (Vol. ASSSP-34, 6, pp. 1666-1667, 1986.
- [20] Shiozaki A. A model of distributed type associated memory with quantized Hadamard transform // *Biol.Cybern.* 38, 1, pp. 19-22, 1980.
- [21] Vose M., Wright A. The simple genetic algorithm and the Walsh transform // *Journal of evolutionary computation*, 6(3), pp. 253-274, 1998.
- [22] Уотермен М.С., редактор. Математические методы для анализа последовательностей ДНК. М., Мир, 1999.
- [23] Лайонс Р. Цифровая обработка сигналов. М., Бином, 2006.
- [24] Jenuwein Th., Allis C.D. Translating the Histone Code. *Science*. Vol. 293, pp. 1074-1080, 2001.
- [25] Шредингер Э. Что такое жизнь? С точки зрения физики. М., Наука, 1972.

# HYPERCOMPLEX NUMBERS AND THE ALGEBRAIC SYSTEM OF GENETIC ALPHABETS. ELEMENTS OF ALGEBRAIC BIOLOGY

S.V. Petoukhov

*Mechanical Engineering Institute, RAS, Moscow, Russia*

spetoukhov@gmail.com

This article presents some results of investigation of the multi-level system of molecular-genetic alphabets by means of matrix methods from theory of noise-immunity coding. These studies have revealed links of the system of alphabets with some systems of hypercomplex numbers (Hamilton quaternions and Cockle split-quaternions and their extensions), Kronecker families of matrices, orthogonal systems of Rademacher functions and Walsh functions, Hadamard matrices etc. Structural parallels are shown between molecular-genetic alphabets and a system of inheritance of traits in holistic organisms, which obeys the Mendel laws and which is reflected in genetic Punnett squares. The system of molecular-genetic alphabets, common to all living organisms, possesses algebraic properties which lead to a new — algebraic — way of cognition of living matter. This cognition is related with development of algebraic biology associated with hypercomplex numbers. Living matter, providing the transmission of hereditary information in the chain of generations, is presented as information entity deeply algebraic in its nature.

**Key Words:** genetic alphabets, hypercomplex numbers, Rademacher functions, Walsh functions, Hadamard matrices, cyclic groups, Punnett squares, algebraic biology.