Нахождение точного частного $Q = \left\lfloor \frac{a}{N} \right\rfloor$ является достаточно сложной проблемой с вычислительной точки зрения, если разрядность чисел, участвующих в операциях, существенно превышает разрядность процессора. Однако в процессе вычисления модулярной экспоненты можно использовать любой вычет из подходящего класса. Окончательный результат приводится к наименьшему вычету после того, как будет выполнено последнее умножение. Например, вместо наименьших вычетов $c = a \cdot b \bmod N$ в процессе вычислений можно использовать вычеты, не являющиеся наименьшими (неприведенные)

$$c = a \cdot b \bmod N + n \cdot N, n \in \{0,1,2\}, \quad (6)$$

вычислить которые значительно легче, чем полностью приведенные к наименьшему значению.

Если $2^{n-3} < N < 2^{n-2}$, то для любого $x < 2^{2n}$ приближение к частному вычисляется следующим образом

$$Q = \left\lfloor \left\lfloor x \cdot 2^{-(n-3)} \right\rfloor \left\lfloor \frac{2^{2n}}{N} \right\rfloor \cdot 2^{-(n+3)} \right\rfloor. \quad (7)$$

Если ввести символические операции $w[i]$ и $[i]w$, означающие взятие i младших и старших бит двоичного представления числа w соответственно, а также записать $\left\lfloor \frac{2^{2k+4}}{N} \right\rfloor$ как $\tilde{N}$, то алгоритм нахождения неприведенного вычета можно представить как

$$\begin{array}{lll} x[2n] & \leftarrow & a[n] \cdot b[n] \\ Q[2n+6] & \leftarrow & \tilde{N}[n+3] \cdot [n+3]x \\ y[n] & \leftarrow & x[n+3] - [n+3]Q \cdot N[n] \end{array} \quad (8)$$

Данный алгоритм в большей степени ориентирован на аппаратную реализацию. В этом случае операции $w[i]$ и $[i]w$ выполняются тривиально, таким образом вычислительные затраты на приведение x в диапазон $[0 \ldots 2^n]$ составят два длинных умножения и одно длинное сложение. В случае программной реализации алгоритма на универсальной архитектуре потребуются дополнительные затраты на получение $n+3$ старших бит чисел x и Q.

Если n кратно 8, то выделение $n+3$ старших бит может быть выполнено посредством сдвига числа вправо на 3 бита. Но в связи с тем, что система команд универсальных процессоров не позволяет быстро выполнять сдвиги чисел произвольной длины на произвольное число бит, использование алгоритма в приведенном виде на такой архитектуре нецелесообразно. В частности, при использовании процессоров типа x86 сдвиг числа длиной m слов на 3 бита потребует m сдвигов двойной точности, столько же команд маскирования результатов сдвига, и объединения по «или» в выходном результате без учета чтения и записи в память. Однако данный алгоритм может быть изменен таким образом, чтобы обойтись без сдвигов вообще.

Рассмотрим соотношение (7) для аппроксимации частного. Выражение $\left\lfloor \frac{2^{2n}}{N} \right\rfloor$ определяет количество удерживаемых значащих разрядов двоичного представления рациональной дроби $\frac{1}{N}$

$$\overbrace{\underbrace{00\ldots01}_{n-2} xx\ldots x}^{2n} . \varepsilon\varepsilon\ldots,$$

которое равно $n+3$. Если n кратно длине машинного слова, для того чтобы удержать и обработать дополнительные 3 бита на универсальной архитектуре потребуется целое слово. Поэтому мы можем увеличить точность двоичного представления рациональной дроби до $n+8$ бит, сохранив при этом как производительность алгоритма, так и количество используемой памяти

$$Q = \left\lfloor \left\lfloor x \cdot 2^{-(n-8)} \right\rfloor \left\lfloor \frac{2^{2n}}{N} \right\rfloor \cdot 2^{-(n+8)} \right\rfloor. \quad (9)$$

Алгоритм, соответствующий (9), имеет вид

$$\begin{array}{lll} x[2n] & \leftarrow & a[n] \cdot b[n] \\ Q[2n+16] & \leftarrow & \tilde{N}[n+8] \cdot [n+8]x \\ y[n] & \leftarrow & x[n+8] - [n+8]Q \cdot N[n] \end{array} \quad (10)$$

В отличие от (8) реализация алгоритма (10) для универсального процессора и побайтно адресуемой памяти не требует выполнения сдвигов, поскольку результат может быть достигнут чтением по другому адресу. Поскольку N является модулем криптосистемы и, в связи с этим, используется многократно, значение выражения $\left\lfloor \frac{2^{2n}}{N} \right\rfloor$ вычисляется один раз и сохраняется как параметр криптосистемы.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. R. L. Rivest, A. Shamir and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems // Communications of the ACM, - Vol.21, - N2, - Feb 1978, - P.120-126.
2. Diffie W., Hellman M. New directions in cryptography. // IEEE Trans. on Information Theory, - Vol.IT-22, - 6 November 1976. - P.644-654.
3. T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms // IEEE Trans. Information Theory, - Vol.IT-31, - 1985, - P.469-472.
4. Nat'l Inst. of Standards and Technology (NIST), FIPS Publication 186: Digital Signature Standard. May 19, - 1994.
5. И. М. Виноградов. Основы теории чисел. - М.: Наука, 1981, - 176с.
6. Л. П. Поденок, Р. Х. Садыхов. Быстрый алгоритм экспоненциального преобразования в системах защиты информации / Тезисы доклада. Труды научно-технической конференции Брестского политехнического института. – 1996. - С.102-105.

Материал поступил в редакцию 22.01.08

PODENOK L.P. Fast algorithms exponention of transformations for crypto-systems with an open key

The method is submitted and the fast algorithm of calculation the module exhibitors based on representation of a parameter as additive-subtraction circuits allowing in average to reduce number of multidigit multiplication in 2 times. The fast method of calculation not of the completely given deductions is submitted.

УДК 577.21

Кузавко Ю.А.

О МАТЕМАТИЧЕСКОЙ ДОПУСТИМОСТИ НЕЕДИНСТВЕННОСТИ СТАНДАРТНОГО, НЕСТАНДАРТНЫХ И ИСКУССТВЕННЫХ КОДОВ ГЕНОМА ЧЕЛОВЕКА

Введение. В молекулярной генетике доказано, что соматическая клетка организма (растения) содержит полную программу воспроизведения организма (растения). Для растений такое явление давно наблюдается в природе, реализовано в экспериментах in vitro. В 2003 г. были завершены многолетние международные экспериментальные исследования кода генома человека, содержащего

3.2 млрд. пар нуклеотидов. При этом был определен код генома не конкретного человека, а нескольких добровольцев-доноров из Буффало [1,2]. Современная автоматическая биомолекулярная аппаратура американской фирмы Celera Corporation позволяет определить геном конкретного человека в течение менее месяца при стоимости работ в 10 млн. долларов. Таким образом, были определены геномы Нобелевского лауреата Уотсона – научного руководителя международной программы «Геном человека», открывшего совместно с Криком в 1953 г. ДНК [3,4] и Нобелевского лауреата Вентера - президента фирмы Celera Corporation. Данные по геномам находятся в свободном Интернет-доступе [1, 5, 6]. Для полной записи генетической информации достаточно одного DVD-диска. Доказано, что только 1.5% генетической информации является полезной, т.е. 48 Мбайт. Количество генов у человека примерно 50 000, т.е. 166 Кбайт. Поставлена задача быстрого определения – за 10 дней для 1000 пациентов с премией в 10 млн. долларов.

1. Генетические коды. К настоящему времени надежно установлена структура 4-х природных генетических кодов: стандартного K^0, нестандартных $K^{1\text{-}3}$, для которых приведены лишь переосмысленные кодонные семейства (таблица 1). Все кодоны содержатся в ДНК, открытой Уотсоном и Криком в 1953 г. [3]. Основной код K^0 будем записывать как $K^0(x_1, x_2, x_3, x_4, x_6)=K^o(2, 9, 1, 5, 3)$. Позиция x_i соответствует n_i кодонам, а индекс i=1,2,3,4,6 означает вырождение состояния относительно кодируемой, одной из 20 аминокислот. Алфавит нуклеотидов является 2-битовым, т.е. реализуется 4-чная позиционная система счисления (ССч) или, точнее, генетический алфавит из четырех букв: A (аденин), G (гуанин), T (тимин), C (цитозин). Аминокислоты генома человека кодируются триплетами нуклеотидов - кодонами. Информационная емкость триплета 4^3=64, т.е. является избыточной для синтеза 20 аминокислот. При этом из них 3 кодона являются терминаторными, т.е. контролирующими сборку мРНК и соответствующих аминокислот. Нестандартные коды K^1-K^3 записывают как $K^1(x_1, x_2, x_3, x_4, x_6)$ и $K^{2,3}(x_1, x_2, x_3, x_4, x_6, x_8)$.

Очевидно, что код K_0 помимо основного представления $K^0(2, 9, 1, 5, 3)$ может иметь еще четыре математических представления: $K_{1,11,0,5,3}$, $K_{3,7,2,5,3}$, $K_{4,5,3,5,3}$ и $K_{5,3,4,5,3}$ в подсемействе преобразований $K_{i,j,\kappa,5,3}$. Два из возможных представлений приведены в таблице (два последних столбца). Подсемейство $K_{i,j,\kappa,5,3}$ включено в семейство основных кодов $K^0_{i,j,k,l,m}$, для которого выполняются следующие уравнения:

$$x_1+2x_2+3x_2+4x_4+6x_6=64\text{-}3=61, \quad (1)$$

$$x_1+x_2+x_2+x_4+x_6=20. \quad (2)$$

Коэффициенты перед x_i в (1) равны степени вырождения мультиплекса кодонов t=3 – количество терминаторных кодонов. Выражение (2) определяет количество кодируемых аминокислот в ядре клетки человека.

Семейство K_1 содержит четыре терминаторных кодона: TAX, AGX (X=A,G), а невырожденное состояние (X_1=0) отсутствует. Семейство $K^{2,3}$ содержит два терминаторных кодона TAX, а также восьмикратно вырожденные состояния: TCN, AGN (N=A,G,T,C). В K^2 также отсутствует невырожденное состояние (X_1=0), а в K^3 имеет место только единственное состояние (X_1=1). Не представляет сложностей для $K^{1\text{-}3}$ записать уравнения аналогичные (1) и (2) для подсчета триплетов в i- состоянии.

Теоретически с точки зрения формальной математики могут существовать другие семейства $K^{4,5,\ldots}$ ($x_1, x_2, x_3, x_4, x_6, \ldots, x_8, \ldots, x_{16}, \ldots, x_{32}$) с различными значениями количества терминаторов t. Программно они все могут быть определены. Пятикратное вырождение состояния x_5 не наблюдалось экспериментально до сих пор как для основного кода K^0, так и для нестандартных кодов $K^{1\text{-}3}$. Возможно это связано с правилами отбора по химическим связям и пространственной симметрии нуклеиновых молекул. В кристаллографии ось симметрии 5-го порядка запрещена: нельзя плотно упаковать плоскость правильными пятиугольниками. Правила отбора могут уменьшить количество элементов в семействе K^0. Вопрос о единственной основной кодировке генома человека может быть подставлен под сомнение. Еще в 1943 г. Шредингер задавался вопросом о новой физике, необходимой для объяснения происхождения жизни на Земле [7].

2. Теорема о неединственности кодировки. Если кодировка K_0 с учетом установленных правил отбора не единственна и будет экспериментально подтверждена, то пятнадцатилетние исследования по расшифровке генома человека не принесли ожидаемого результата, т.е. методической основы для технологического создания вакцин против летальных заболеваний человека: СПИД, рак и т.д. Действительно, в течение четырех лет такие вакцины не созданы. Возникло новое вирусное заболевание – птичий грипп, быстро мутирующий и смертельно опасный для людей. Косвенным подтверждением не единственности кодировки K^0 явилось открытие нестандартных кодировок $K^{1\text{-}3}$ в митохондриях клетки.

Теперь сформулируем математическую теорему об основной кодировке K^0.

Теорема. Преобразование (кодировка) K^0 триплетами кодонов матричной РНК и аминокислот не является единственной и содержит 5+7+1+5+1=19 вариантов.

Доказательство следует из уравнений (1), (2) методом комбинаторного перебора. Легко формулируются аналогичные теоремы для кодировок $K^{1\text{-}3,\ 4\ldots}$. Количество вариантов подсчитывается программно.

В митохондриях клетки человека также реализуется синтез десяти аминокислот. Для этого достаточно дуплетной организации нуклеотидов (4^2=16). Для простейших микроорганизмов – ретровирусов, не содержащих иногда ДНК, имеет место монокодирование, т.е. только четыре нуклеотида кодируют аминокислоты. Экспериментально наблюдалось кодирование не более 3 аминокислот. Процесс происходит только в присутствии клетки – хозяина. Репликация ретровирусов очень опасна для человека, обуславливая онкологические заболевания, которым подвержено 200 млн. людей. Установление истинной картины транскрипции, трансляции и репликации ретровирусов должно дать ключ к выяснению механизмов их ингибирования в клетках высших эукариот, что является многообещающим направлением терапии раковых заболеваний.

Проведем некоторые аналогии биологического кодирования с преобразованием чисел из одной позиционной системы счисления в другую, преобразованием алфавита в лингвистике.

Любое число однозначно может быть представлено в любой позиционной n-ной ССч. Исторически сложилась десятичная ССч. В компьютерах вычисления осуществляются в двоичной ССч, которая обеспечивает максимальную устойчивость вычислительных систем. Программно используется 8-чная и 16-чная ССч. Преобразование числа из одной позиционной ССч в другую может выполняться через промежуточную ССч. Преимуществом обладает двоичная ССч, при этом при поразрядном переводе избыток разряда переводится в следующий разряд.

В формальной лингвистике существует множество алфавитов A_i с разным буквенным наполнением. Прямой переход $A_1\{a_1\}$ -> $A_2\{a_2\}$ не всегда может быть однозначным, обладает вырождением и отсутствием поразрядного переноса. Генетическое кодирование наиболее близко к лингвистическому преобразованию алфавитов.

Приведем определения понижающего и повышающего, эквивалентного преобразований (кодирований) алфавитов.

Определение 1. Преобразование $A_1\{a_1\}$ -> $A_2\{a_2\}$ является понижающим, если мощности алфавитов соотносятся как $N_{a1} > M_{a2}$.

Определение 2. Преобразование $A_2\{a_2\}$ -> $A_1\{a_1\}$ является повышающим, если мощности алфавитов соотносятся как $M_{a2} < N_{a1}$.

Определение 3. Преобразование $A_1\{a_1\}$ -> $A_2\{a_2\}$ является эквивалентным (тривиальным), если мощности алфавитов равны: $N_{a1} = M_{a2}$.

Таблица 1. *Структура для 4-х природных и некоторых генетических кодов: для стандартного K^0, нестандартных $K^{1\text{-}3}$, для которых приведены лишь переосмысленные кодонные семейства (см. ниже) и искусственных $K^{4,\dots}$ кодов*

Аминокислота	K^0	K^1	K^2	K^3	$K^4{}_{3,7,2,5,3}$	$K^5{}_{4,5,3,5,3}$
1 Met	(1) ATG	(2) ATX	(2) ATX		(1) ATG	(1) ATG
2 Trp	(1) TGG	(2) TGX	(2) TGX	(2) TGX	(1) TGG	(1) TGG
3 Phe	(2) TTY				(1)	(1)
4 Tyr	(2) TAY				(2) TAY	(1)
5 His	(2) CAY				(2) CAY	(2) CAY
6 Asn	(2) AAY			(3) AAM	(2) AAY	(2) AAY
7 Asp	(2) GAY				(2) GAY	(2) GAY
8 Cys	(2) TGY				(2) TGY	(2) TGY
9 Gln	(2) CAX				(2) CAX	(2) CAX
10 Lys	(2) AAX			(1) AAG	(2) AAX	(3) AAM
11 Gly	(2) GAX				(3) ATM	(3) ATM
12 Ile	(3) ATM	(2) ATY	(2) ATY		(3)	(3)
13 Val	(4) GTN				(4) GTN	(4) GTN
14 Pro	(4) CCN				(4) CCN	(4) CCN
15 Thr	(4) CAN				(4) CAN	(4) CAN
16 Ala	(4) GCN				(4) GCN	(4) GCN
17 Gly	(4) GGN				(4) GGN	(4) GGN
18 Ser	(6) TCN, AGY		(8) TCN, AGN	(8) TCN, AGN	(6) TCN, AGY	(6) TCN, AGY
19 Leu	(6) CTN, TTX	(4) CGN			(6) CTN, TTX	(6) CTN, TTX
20 Arg ter	(3) TAX, TGA	(4) TAX, AGX	(4) CGN (2) TAX	(4) CGN (2) TAX	(3) TAX, TGA	(3) TAX, TGA

Примечание. При записи 20-и аминокислот были использованы стандартные трехбуквенные сокращения. Для стандартного кода K^0 указано число кодонов-синонимов (в скобках) и их трехбуквенные представления. Обозначения: X: A,G; Y: T,C; .M: T,C,A; N: A,G,T,C. В последней строке приводятся три терминаторных кодона – ter, каждый из которых обозначает останов синтеза белка

Лема. Если прямое кодирование является понижающим (повышающим), то обратное кодирование – повышающим (понижающим).

Очевидно, что эквивалентные преобразования всегда однозначны и невырождены. Для передачи генетической информации они несущественны. Точный англоязычный перевод русской прозы не несет никакой дополнительной информационной ценности по сравнению с оригиналом. Обратное к понижающему кодированию преобразование является повышающим. При его реализации теряется информация. Это основная причина, по которой белки не могут синтезировать РНК и следовательно ДНК. В принципе повышающее кодирование может быть реализовано. Как физически организовать преобразование одной буквы исходного алфавита в две буквы конечного алфавита? Такое преобразование может быть только квантовым. До настоящего времени в молекулярных нанотехнологиях считалось [8], что квантовые процессы в живой природе не происходят. Исключением явилось недавнее сообщение, указавшее на квантовый характер механизма обоняния человека. Выдвинем гипотезу: существует вероятность транскрипции, трансляции и репликации белками фрагментов мРНК, обусловленная квантовым запутыванием состояний активаторов белка.

Заключение. Сформулируем две основные гипотезы генетического кодирования и декодирования, требующие дальнейшего доказательства, но логичные по своей форме.

Гипотеза генетического кодирования

Преобразование двух генетических алфавитов соответственно мощностями M и N является понижающим и вырожденным. Избыточность кодирования $K=M-N=k_1+k_2+\dots+k_n+t$ организуется мультиплексорно по n-вырождениям с весами $k_1, k_2, \dots, k_n$. Число t представляет количество терминаторных букв (кодонов) для управления единичной кодировкой. Существуют правила отбора кодирований, исключающие возможные формально математические преобразования.

Квантовая гипотеза генетического декодирования

Преобразование генетических алфавитов соответственно мощностями M и N с повышением классическими механизмами запрещено. Квантовый механизм декодирования алфавитов допускается в пределах времен декогерентности состояний квантового мультиплета.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Сайт AnimatLab: http://animatlab.lip6.fr/index.en.html
2. Козлов Н.Н. Математический анализ генетических кодов. Математическая биология и биоинформатика. 2006. Т. 1, № 1. - С. 70-96.
3. Watson J.D., Crick F.H.L. A structure for Deoxeribose Nucleic Acid. Nature. 1953. V. 171. P. 737-738.
4. Уотсон Д. Двойная спираль. Воспоминания об открытии структуры ДНК. - М.: Мир, 1969. - 152 с.
5. Козлов Н.Н. О востребованности каждого из 64 кодонов в генетических перекрытиях. Доклады академии наук. 1999. Т. 367, № 4. - С. 544-547.
6. Козлов Н.Н. Теорема для генетического кода. Доклады академии наук. 2002. Т. 382, № 5. - С. 593-597.
7. Шредингер Э. Что такое жизнь? С точки зрения физика. - М.: Атомиздат, 1972. - 88 с.
8. Редько В.Г. Эволюционная кибернетика. - М.: Наука, 2001. - 156 с.

Материал поступил в редакцию 27.01.08

KUZAVKO J.A. About a mathematical admissibility not uniform standard, non-standard and artificial codes genom of the man

From the point of view of formal mathematics is shown, that the coding DNK matrix PNK and appropriate to her aminoacids can occur under the different scripts. 4 natural genetic codes are complemented: standard K_0, non-standard codes $K_{1\text{-}3}$, both in family by them formed, and in artificial K_4, …. The rules of selection of the scripts of coding are proved and the basic theorem about not uniform of the coding K_0 threefold nucleothid mRNK and appropriate to her aminoacids is proved. The coding in mithochondria of a crate as double coding, in onco retroviruses - as monocoding is discussed. The equation of definition of natural and artificial transformations of the genetic information for calculation of quantity of genetic codes is written down. The hypotheses of direct genetic coding and return (quantum) genetic decoding are given.