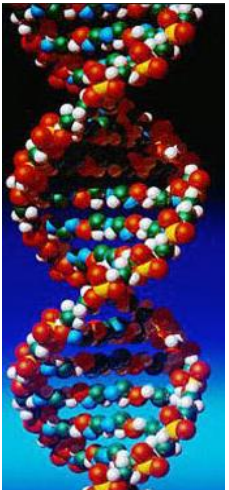
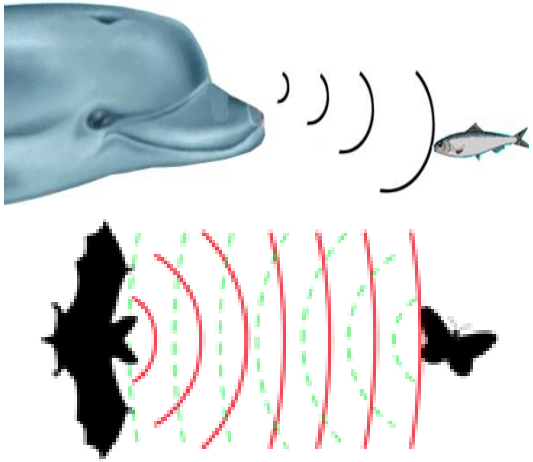
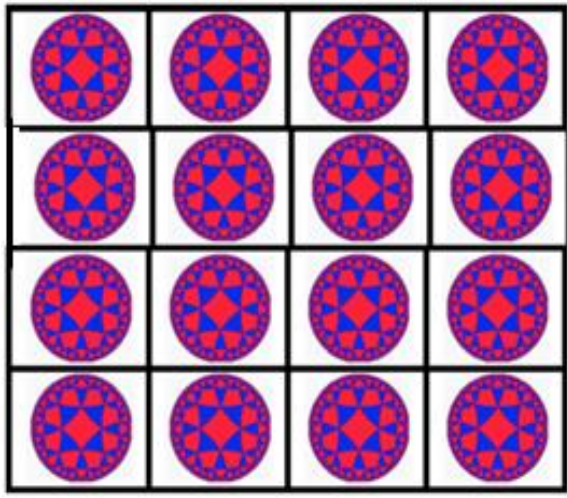
МАТРИЧНАЯ ГЕНЕТИКА И АЛГЕБРАИЧЕСКАЯ БИОЛОГИЯ (MATRIX GENETICS AND ALGEBRAIC BIOLOGY)

С.В. Петухов (S.V. Petoukhov)

Институт машиноведения им. А.А. Благонравова РАН
http://petoukhov.com/, spetoukhov@gmail.com

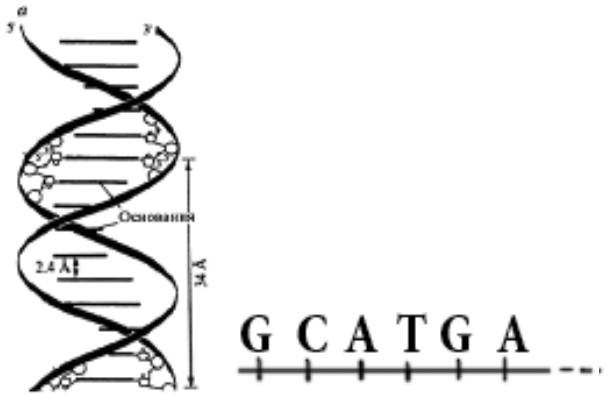


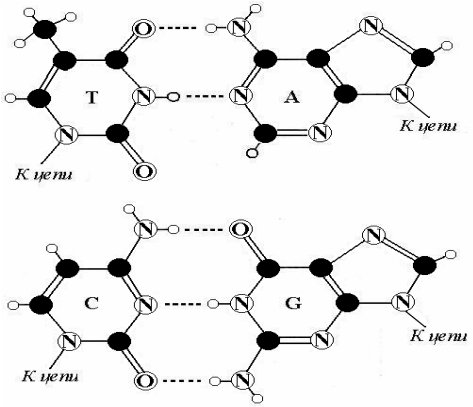



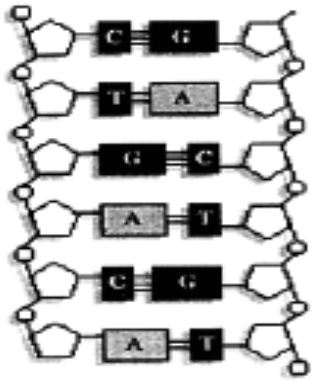
Живые организмы наделены врожденными интеллектуальными способностями для поиска пищи, спасения от хищников, спаривания, строительства (паутина пауков, соты пчел, сложные гнезда птиц-ткачей и пр.). При создании систем искусственного интеллекта следует **подсматривать у живой природы алгоритмы и универсальные структуры генетического наследования таких способностей.** Доклад посвящен некоторым результатам такого подсматривания через призму алгебры.

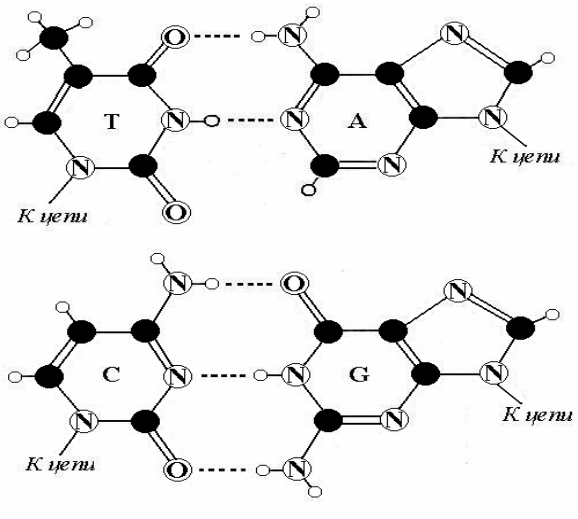
Бинарные-оппозиции в алфавитах ДНК и матрицы

Генетическая информация во всех организмах записывается единообразно в виде длинных текстов из четырех молекулярных «букв» (нуклеотидов). В молекулах ДНК такими буквами выступают аденин A, цитозин C, гуанин G и тимин T (в РНК вместо тимина - урацил U). Помимо алфавита 4 нуклеотидов в ДНК имеются алфавиты 16 дуплетов, 64 триплетов и пр., построенных на комбинациях 4 нуклеотидов.









Алфавит 4 нуклеотидов ДНК является носителем системы бинарно-оппозиционных признаков:

1) два из этих нуклеотидов являются пуринами (A и G), а два другие (C и T) – пиримидинами. С позиций этих оппозиционных индикаторов **C = T = 1, A = G = 0**;

2) два из этих нуклеотидов являются кето-молекулами (T и G), а два другие (C и A) – амино-молекулами, что дает представление **C = A = 1, T = G = 0**.

В силу этого ДНК-алфавиты 4 букв, 16 дуплетов и 64 триплетов представляются в форме квадратных таблиц, столбцы которых нумеруются бинарными индикаторами «пиримидин или пурин» (**C=T=0, A=G=1**), а строки – бинарными индикаторами «амино или кето» (**C=A=0, T=G=1**). В таких таблицах все буквы, дуплеты и триплеты автоматически занимают строго индивидуальное место.

	0	1
0	C	A
1	T	G

;

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

	0	1
0	C	A
1	T	G

;

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

Эти 3 таблицы являются не просто таблицами, но образуют единое **тензорное семейство матриц** (на основе операции конкатенации): вторая и третья тензорная (кронекеровская) степень (2*2)-матрицы [C, A; T, G] автоматически дают эту (4*4)-матрицу алфавита 16 дуплетов и эту (8*8)-матрицу алфавита 64 триплетов, которые кодируют 20 аминокислот в белках.

Напомним, что тензорное произведение матриц играет важную роль в квантовой информатике и квантовой механике многокомпонентных систем: **пространство состояния многокомпонентной системы является тензорным произведением пространств состояний ее компонент**.

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

Как же в этой матрице 64 триплетов расположены кодируемые ими 20 аминокислот и стоп-сигналы белкового синтеза? Это расположение нельзя предсказать, поскольку аминокислоты и нуклеотиды A, C, G, T совершенно различны по строению. Число вариантов расположения аминокислот с их какими-то повторениями для заполнения всей (8*8)-матрицы необозримо: **>>10^{100}** (для сравнения в физике время существования Вселенной оценивается в 10^{17} секунд). **Будет ли это расположение аминокислот хаотичным или вдруг окажется закономерно симметричным?**

Оказывается, что из океана возможностей природа выбрала алгебраически закономерный вариант повторения и расположения аминокислот и стоп-сигналов в этой матрице 64 триплетов (показан случай генетического кода митохондрий позвоночных):

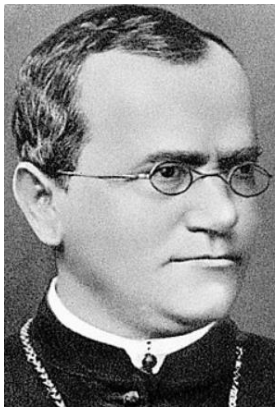
	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

CCC Pro	CCA Pro	CAC His	CAA Gln	ACC Thr	ACA Thr	AAC Asn	AAA Lys
CCT Pro	CCG Pro	CAT His	CAG Gln	ACT Thr	ACG Thr	AAT Asn	AAG Lys
CTC Leu	CTA Leu	CGC Arg	CGA Arg	ATC Ile	ATA Met	AGC Ser	AGA Stop
CTT Leu	CTG Leu	CGT Arg	CGG Arg	ATT Ile	ATG Met	AGT Ser	AGG Stop
TCC Ser	TCA Ser	TAC Tyr	TAA Stop	GCC Ala	GCA Ala	GAC Asp	GAA Glu
TCT Ser	TCG Ser	TAT Tyr	TAG Stop	GCT Ala	GCG Ala	GAT Asp	GAG Glu
TTC Phe	TTA Leu	TGC Cys	TGA Trp	GTC Val	GTA Val	GGC Gly	GGA Gly
TTT Phe	TTG Leu	TGT Cys	TGG Trp	GTT Val	GTG Val	GGT Gly	GGG Gly

Показанная справа матрица кодируемых аминокислот и стоп-сигналов состоит из пар соседних строк, идентичных по составу аминокислот и стоп-кодонов. Эта тензорно-матричная закономерность является одним из свидетельств глубокой связи генетического кодирования с формализмами квантовой информатики.

Дуализм «вероятность-детерминизм» в генетике

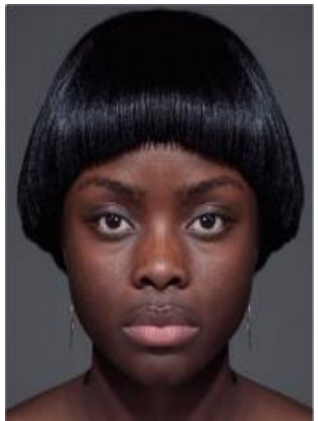
Отдельные молекулы взаимодействуют в клетках стохастическим образом. В живых телах все строится на стохастике. Даже генетически идентичные клетки одной ткани имеют разные уровни экспрессии белков, размеры и пр. Но при этой стохастике в «малом» макро-телесные признаки наследуются от родителей как детерминированные. Генетика как наука началась с открытия Менделем стохастических правил наследования признаков при скрещивании организмов.



Гр. Мендель (1822-1884)

Согласно закону Менделя независимого наследования признаков, информация с уровня молекул ДНК диктует макроструктуры живых тел по многим независимым каналам, несмотря на сильные помехи. Так, цвета волос, глаз и кожи наследуются независимо друг от друга.

Соответственно, **каждый организм является машиной многоканального помехоустойчивого кодирования на основе принципов стохастики.**



ОТЛИЧИЕ ЖИВОГО ОТ НЕЖИВОГО

Создатели квантовой механики П.Йордан и (позже) Э.Шредингер указывали на **ключевое отличие живых тел от неодушевленных**: **неодушевленные объекты управляются средним случайным движением их миллионов частиц и движение отдельных частиц не существенно для целого; напротив, в живом организме избранные – генетические - молекулы обладают диктаторским влиянием на весь организм** за счет квантового усиления (см. историю «квантовой биологии», 2018).

Йордан утверждал, что упущенные наукой законы **жизни являются законами вероятностей квантового мира.**

Pascual Jordan

Erwin Schrödinger

В поиске этих упущенных законов жизни обратимся к изучению стохастических закономерностей в геномных ДНК. В двойной спирали ДНК комплементарные пары нуклеотидов A-T и C-G связаны соответственно 2 и 3 водородными связями. Поэтому вся последовательность нуклеотидов нити ДНК может быть представлена в виде последовательности двоек и троек типа 23323322… Для конкретики сосредоточимся на анализе ДНК первой хромосомы человека из общедоступного GenBank. Она содержит текст из около 250 миллионов нуклеотидов (показан ее фрагмент).

```
TCTGACCTGAGGAGAACTGTGCTCCGCCTTCAGAGTACCACCGAAATCTGTGCAGAGGACAACGCAGCTC
CGCCCTCGCGGTGCTCTCCGGGTCTGTGCTGAGGAGAACGCAACTCCGCCGTTGCAAAGGCGCGCCGCGC
CGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGC
AGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGC
CGCGCCGGCGCAGGCGCAGACACATGCTAGCGCGTCGGGGTGGAGGCGTGGCGCAGGCGCAGAGAGGCGC
GCCGCGCCGGCGCAGGCGCAGAGACACATGCTACCGCGTCCAGGGGTGGAGGCGTGGCGCAGGCGCAGAG
AGGCGCACCGCGCCGGCGCAGGCGCAGAGACACATGCTAGCGCGTCCAGGGGTGGAGGCGTGGCGCAGGC
GCAGAGACGCAAGCCTACGGGCGGGGGTTGGGGGGGCGTGTGTTGCAGGAGCAAAGTCGCACGGCGCCGG
GCTGGGGCGGGGGGAGGGTGGCGCCGTGCACGCGCAGAAACTCACGTCACGGTGGCGCGGCGCAGAGACG
```

Дихотомии вероятностей в стохастической организации водородных связей 2 и 3 в геномных ДНК.

```
TCTGACCTGAGGAGAACTGTGCTCCGCCTTCAGAGTACCACCGAAATCTGTGCAGAGGACAACGCAGCTC
CGCCCTCGCGGTGCTCTCCGGGTCTGTGCTGAGGAGAACGCAACTCCGCCGTTGCAAAGGCGCGCCGCGC
CGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGC
AGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGC
CGCGCCGGCGCAGGCGCAGACACATGCTAGCGCGTCGGGGTGGAGGCGTGGCGCAGGCGCAGAGAGGCGC
GCCGCGCCGGCGCAGGCGCAGAGACACATGCTACCGCGTCCAGGGGTGGAGGCGTGGCGCAGGCGCAGAG
AGGCGCACCGCGCCGGCGCAGGCGCAGAGACACATGCTAGCGCGTCCAGGGGTGGAGGCGTGGCGCAGGC
GCAGAGACGCAAGCCTACGGGCGGGGGTTGGGGGGGCGTGTGTTGCAGGAGCAAAGTCGCACGGCGCCGG
GCTGGGGCGGGGGGAGGGTGGCGCCGTGCACGCGCAGAAACTCACGTCACGGTGGCGCGGCGCAGAGACG
```

Представляя нуклеотиды их числами водородных связей A=T=2 и C=G=3, получаем представление этой ДНК в виде бинарной последовательности из около 250 миллионов цифр 2 и 3 типа: 22323323323223…..

Для обнаружения скрытых в этой последовательности стохастических (вероятностных) закономерностей проанализируем ее по авторскому методу «стохастической матрешки», рассматривающей ее устройство как аналог устройства русских матрешек:

Рассмотрим сначала изучаемую бинарную последовательность вида 23322322… как последовательность одиночных цифр 2-3-3-2-2-.., в которой вычисляем процентное содержание цифр 2 и 3, т.е. %2 и %3.

Затем ту же последовательность рассматриваем как последовательность цифровых дуплетов 23-32-23-22… и вычисляем процент каждого из 4 возможных видов дуплетов: %22, %23, %32, %33.

Затем аналогично рассматриваем ту же последовательность как последовательность цифровых триплетов, тетраплетов, пентаплетов, …, вычисляя всякий раз проценты каждого из видов триплетов, тетраплетов и т.д.

Другими словами, мы рассматриваем данный бинарный текст ДНК как набор или матрешку многих параллельных текстов, каждый из которых написан на своем собственном алфавите *n*-плетов.

В результате для ДНК хромосомы человека №1 получаем таблицу процентов *n*-плетов ее водородных связей 2 и 3 в разных *n*-текстовых слоях (проценты даны в долях единицы):

%2	%3
0,582757	0,417243

%22	%23	%32	%33
0,328129	0,254635	0,254622	0,162614

%222	%223	%232	%233	%322	%323	%332	%333
0,200289	0,127765	0,155746	0,098982	0,127812	0,126809	0,098968	0,063630

%2222	%2223	%2232	%2233	%2322	%2323	%2332	%2333
0,12754988	0,07278454	0,08176248	0,0460202	0,08159258	0,07408348	0,05981373	0,039135
%3222	%3223	%3232	%3233	%3322	%3323	%3332	%3333
0,072757	0,055044	0,073902	0,05294	0,046241	0,052734	0,039127	0,024518

%22222	%22223	%22232	%22233	%22322	%22323	%22332	%22333
0,0841722	0,0433358	0,047984	0,0247861	0,0455768	0,0361372	0,0295058	0,016538
%23222	%23223	%23232	%23233	%23322	%23323	%23332	%23333
0,0478955	0,0337308	0,0451478	0,0289408	0,0295923	0,0302517	0,0249346	0,014141
%32222	%32223	%32232	%32233	%32322	%32323	%32332	%32333
0,043374	0,0294667	0,0337435	0,0212295	0,0360132	0,0379457	0,0303232	0,0225501
%33222	%33223	%33232	%33233	%33322	%33323	%33332	%33333
0,0248969	0,0213263	0,0287649	0,0239821	0,0166674	0,02250250	0,01417871	0,01036468

Совершенно неожиданным оказалось, что в этих данных многослойной стохастики геномной ДНК между процентами *n*-плетов водородных связей из семейств с разными значениями ***n*** имеется **высокоточная дихотомическая** связь (при изученных значениях ***n*** = 1, 2, 3, 4, 5): процент любого ***n***-плета практически равен сумме процентов таких двух (***n***+1)-плетов, которые отличаются от него наличием суффиксов 2 и 3. Например:

%2=**0,58276** ≈ %22+%23=**0,58276**;

%3=**0,41724** ≈ %32+%33= **0,41724**;

%22=**0,3281** ≈ %222+%223=**0,3281**;

%32=**0,25462** ≈ %322+%333=**0,25462;**

%23=**0,2546** ≈ %232+%233 ≈ **0,2547**;

%33=**0,1626** ≈ %332+%333=**0,1626**; и т.д.

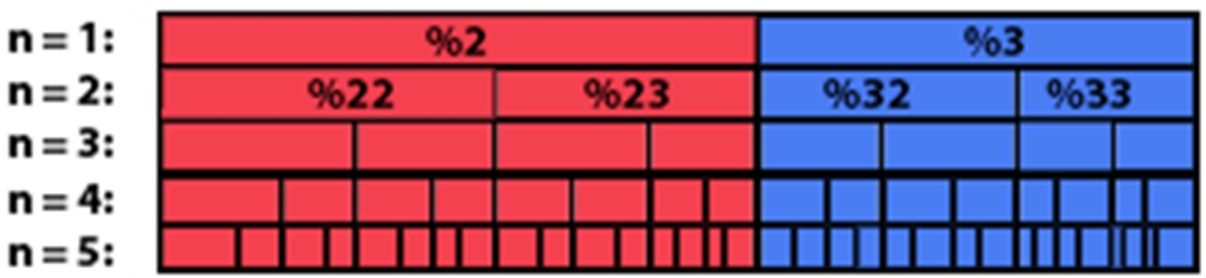
%2 ≈

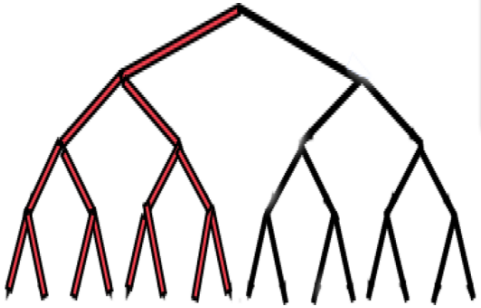
%22 + %23

Точность равенств до 4- или 5-го знака после запятой.

Выполнение этого правила дихотомической взаимосвязи между процентами из соседних семейств n-плетов ведет к существованию фрактального дихотомического дерева вероятностей. В этом дереве каждый член является началом своего собственного дихотомического дерева.

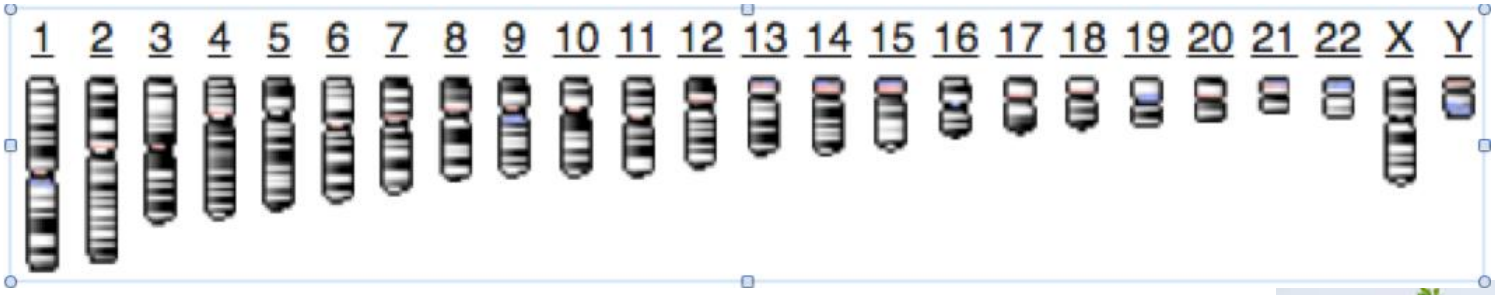
∑ %	Percentage of even H-*n*-plets starting with 2	Percentage of odd H-*n*-plets starting with 3	∑ %
0,58275	%2	%3	0,4172
0,58276	%23+%22	%33+%32	0,4172
0,58278	%233+%232+%223+%222	%333+%332+%323+%322	0,4172
0,58274	%2333+%2332+%2323+%2322 +%2233+%2232+%2223+%2222	%3333+%3332+%3323+%3322 +%3233+ %3232+%3223+%3222	0,4173
0,58267	%23333+%23332+%23323+%23322 +%23233+%23232+%23223+%23222 +%22333+%23332+%22323+%22322 +%22233+%22232+%22223+%22222	%33333+%33332+%33323+%33322 +%33233+%33232+%33223+%33222 +%32333+%32332+%32323+%32322 +%32233+%32232+%32223+%32222	0,4173



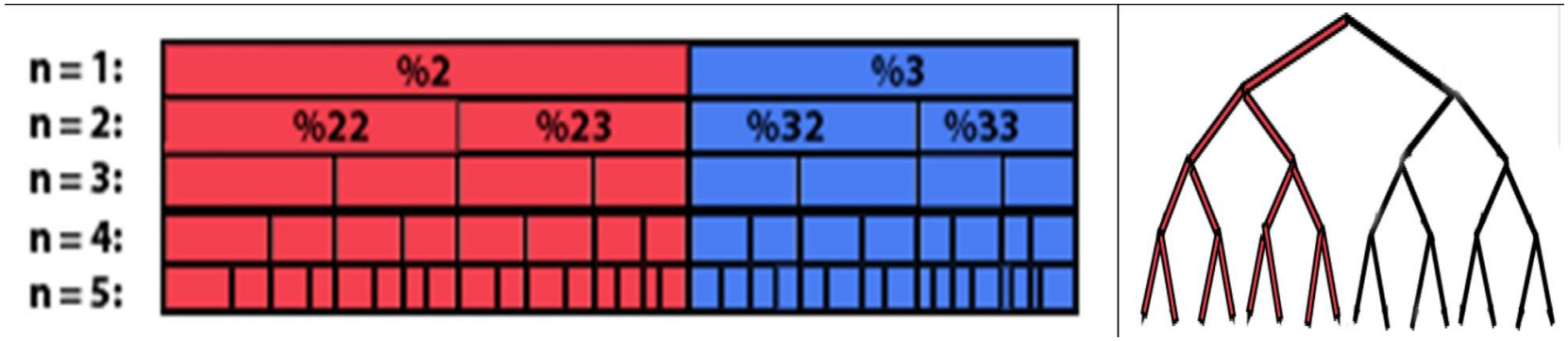


ГЕНОМЫ ЭУКАРИОТ И ПРОКАРИОТ

Это и другие обнаруженные правила стохастиче-ской организации последовательностей водородных связей 2 и 3 справедливы для всех уже исследованных в нашей лаборатории геномных ДНК:
- 1) всех 24 хромосом человека; 2) всех хромосом дрозофилы, мыши, червя, многих растений;
3) 19 геномов бактерий и архей;
4) многих экстремофилов, живущих в экстремальных условиях, например, радиации с уровнем в 1000 раз превышающим смертельный для человека.

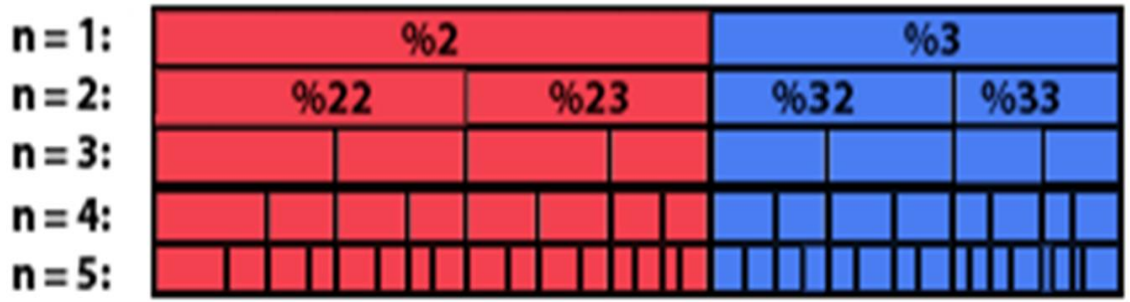


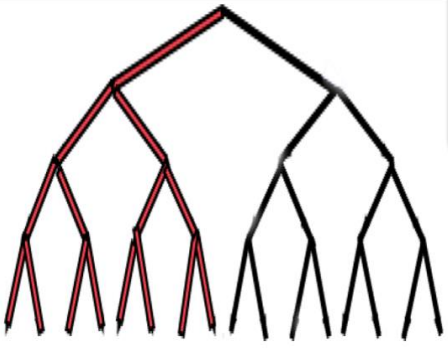
Данные универсальные правила геномных ДНК говорят о существовании нетривиальных алгебраических инвариантов **глобального** геномного характера, которые остаются неизменными на протяжении миллиардов лет эволюции, в ходе которой отмирают миллионы видов организмов и возникают новые (хотя **локально** геномные последовательности изменяются мутациями, прессом естественного отбора и пр.).



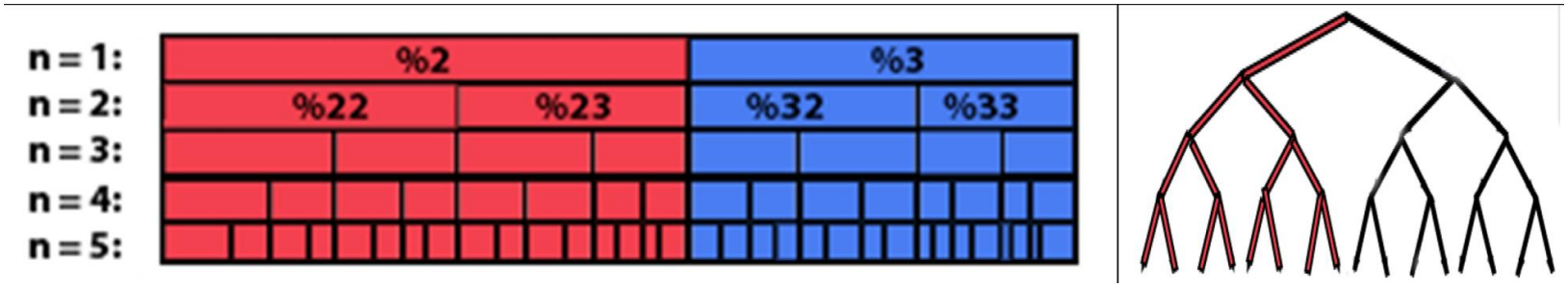
Дихотомии в биологических телах хорошо известны: бронхиальное дерево легких человека с его 23 уровнями дихотомий, дихотомические ветвления в нейронах, ветвления у растений, митоз соматических клеток, ….

- Полученная асимметричная пара из левого и правого фрактальных деревьев вероятностей в геномных ДНК, связанная с четным и нечетным числами 2 и 3, ассоциируется с асимметричными полушариями головного мозга. Сквозная для биологии и культуры тема Инь-Ян (левое и правое, чет и нечет, женское и мужское) систематизирована в известной книге Иванова Вяч.Вс. «Чет и нечет. Асимметрия мозга и знаковых систем». Онкология и многие телесные биоструктуры связаны с фракталами.





Но в геномных ДНК в отличие от телесных конструкций мы встречаемся с принципиально иным типом дихотомии: дихотомиями процентных (вероятностных) характеристик в многослойной стохастике информационных последовательностей ДНК. Обширные дихотомические фрактальные сети вероятностей геномных ДНК – это та информационная почва, из которой произрастают живые тела и генетический интеллект. Вещественные дихотомические структуры живых тел возникают не на пустом месте, а имеют структурные прототипы в закономерной бинарной фрактало-подобной системе вероятностей геномной информатики.

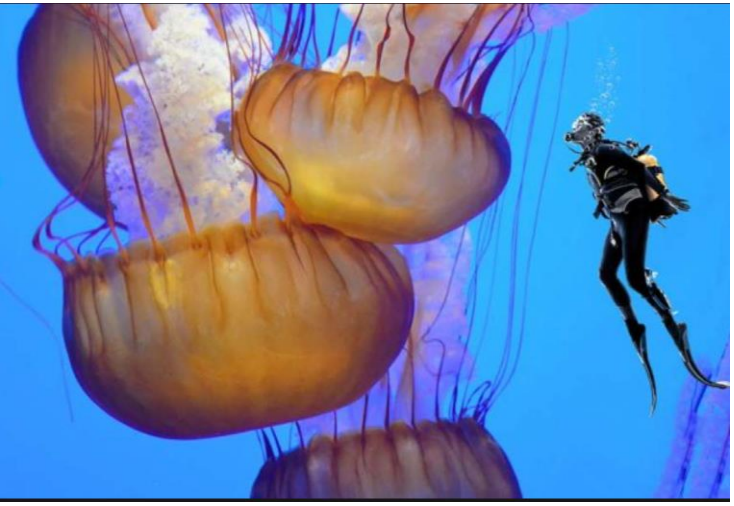


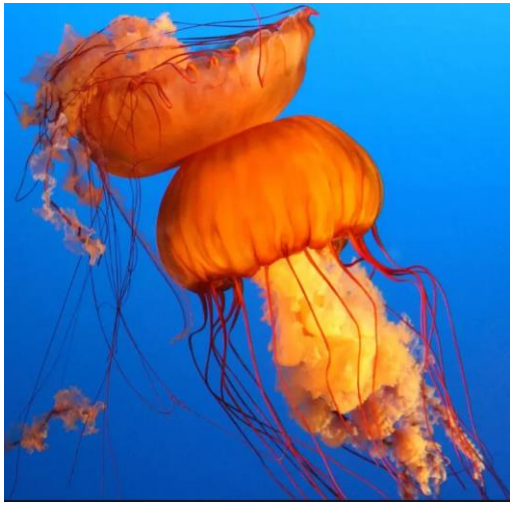
Выявление универсальных стохастических дихотомий в геномах позволяет по-новому осмысливать явления дихотомий в биологических телах. Например, почему на протяжении миллиардов лет жизни на Земли бактерии размножаются дихотомическим делением тела на две половины? На какие структурно-энергетические основы опирается этот «вечный» дихотомический феномен размножения бактерий со сложнейшим процессом дихотомического разделения всей дихотомически организованной генетической информации?

Автор предлагает следующий ответ: **существует скрытый от непосредственного восприятия мир семейств вероятностей, структурированных на основе бинарных оппозиций (типа Инь-Ян).** Именно по образу и подобию бинарно организованных семейств вероятностей этого многомерного мира строятся генетически наследуемые биологические тела. Наши тела являются как бы одеждой, надетой на эти бинарно структурированные семейства вероятностей, выступающие прототипами биологических тел.

Это напоминает ситуацию с человеком-невидимкой из романа Г. Уэльса, невидимая фигура которого проявляется только при ношении им одежды. Это также напоминает древние представления о проявленном и непроявленном мирах и знаменитую аллегорию Платона о мире идей и тенях на стене пещеры, по которым люди, живущие в пещере, могут судить об истинном **скрытом мире идей**.

Лишь кратко отметим, что доложенные универсальные правила геномных ДНК связаны с водородными связями, которые являются – по Л.Полингу - важнейшей для физиологии структурой. Водородные связи определяют конфигурации белков и многие физиологические процессы, а также свойства воды и льда. Медузы на 98% состоят из воды и представляют самый древний вид многоклеточных животных с богатейшим эволюционным многообразием и обилием жизненных функций.





Описанные выше алфавитные матрицы *n*-плетов нуклеотидов представляются как бисимметричные матрицы *n*-плетов водородных связей 2 и 3 (A=T=2, C=G=3) и как матрицы вероятностей этих n-плетов в геномных ДНК. Например, матрица дуплетов:

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

→

	00	01	10	11
00	33	32	23	22
01	32	33	22	23
10	23	22	33	32
11	22	23	32	33

→

	00	01	10	11
00	%33	%32	%23	%22
01	%32	%33	%22	%23
10	%23	%22	%33	%32
11	%22	%23	%32	%33

Эти (2^n*2^n)-матрицы процентов водородных связей имеют собственные числа и **характеристические полиномы**, соответствующие стохастическим особенностям ДНК геномов и генов. **Универсалии алгебраической генетики оказываются сопряженными с алгебраической геометрией**, которая фокусируется на нулях полиномов и является одной из центральных областей современной математики.

Декомпозиции каждой из этих матриц процентов показывают, что она является матричным представлением соответствующего 2^n-мерного гиперболического числа. Например, декомпозиция матрицы процентов 4 дуплетов имеет вид $\%33*e_0+\%32*e_1+\%23*e_2+\%22*e_3$, где набор разреженных матриц e_0, e_1, e_2, e_3 замкнут относительно умножения и определяет известную таблицу умножения алгебры 4-мерных гиперболических чисел (показано):

$$\begin{vmatrix} \%33, \%32, \%23, \%22 \\ \%32, \%33, \%22, \%23 \\ \%23, \%22, \%33, \%32 \\ \%22, \%23, \%32, \%33 \end{vmatrix} = \%33 \begin{vmatrix} 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \\ 0,0,0,1 \end{vmatrix} + \%32 \begin{vmatrix} 0,1,0,0 \\ 1,0,0,0 \\ 0,0,0,1 \\ 0,0,1,0 \end{vmatrix} + \%23 \begin{vmatrix} 0,0,1,0 \\ 0,0,0,1 \\ 1,0,0,0 \\ 0,1,0,0 \end{vmatrix} + \%22 \begin{vmatrix} 0,0,0,1 \\ 0,0,1,0 \\ 0,1,0,0 \\ 1,0,0,0 \end{vmatrix}$$

*	e_0	e_1	e_2	e_3
e_0	e_0	e_1	e_2	e_3
e_1	e_1	e_0	e_3	e_2
e_2	e_2	e_3	e_0	e_1
e_3	e_3	e_2	e_1	e_0

ДНК всех геномов и генов по их стохастическим характеристикам представимы 2^n-мерными гиперболическими числами. На этой основе предлагается единая бинарно-стохастическая паспортизация всех ДНК для их сравнения на базе операций этих алгебр и соответствующего многомерного метрического анализа.

КУБИТЫ И ВЕРОЯТНОСТИ N-ПЛЕТОВ ГЕНОМНЫХ ДНК

Открытие универсальных правил вероятностей в квази-стохастической бинарной организации геномных ДНК привлекло внимание к связи этих правил с кубитами квантовой информатики. Формально можно представить бинарные последовательности водородных связей 2 и 3 как двухуровневую квантово-подобную систему. Состояние ее «генетического» кубита для моноплетов

$$|\psi_1\rangle = \sqrt{\%2}\,|0\rangle + \sqrt{\%3}\,|1\rangle$$

соответствует вектору в 2-мерном гильбертовом пространстве. Здесь $\sqrt{\%2}$ и $\sqrt{\%3}$ являются амплитудами вероятностей состояний $|0\rangle$ и $|1\rangle$, соответствующих состояниям 2 и 3 водородные связи ($\%2 + \%3 = 1$). По аналогии, можно ввести n-кубиты для случаев вероятностей n-плетов водородных связей.

Например, для вероятностей дуплетов имеем 2-кубит:

$$|\psi_2\rangle = \sqrt{\%22}\,|00\rangle + \sqrt{\%23}\,|01\rangle + \sqrt{\%32}\,|10\rangle + \sqrt{\%33}\,|11\rangle$$

Используя феноменологические данные вероятностей *n*-плетов, например, для ДНК первой хромосомы человека, получаем, что тензорный квадрат состояния кубита моноплетов водородных связей не равен состоянию 2-кубита дуплетов водородных связей:

$$|\psi_1\rangle \otimes |\psi_1\rangle \neq |\psi_2\rangle$$

Подобные неравенства в квантовой информатике означают, что состояние многокомпонентной квантовой системы не является сепарабельным чистым состоянием, то есть является запутанным (entanglement) [Nielsen, Chuang, 2010]. По формальной аналогии, можно говорить о запутанном квантово-подобном состоянии многокомпонентных последовательностей геномных ДНК.

- Вектора из амплитуд вероятностей таких «генетических» кубитов (например, $[\sqrt{\%2}, \sqrt{\%3}]$ и $[\sqrt{\%22}, \sqrt{\%23}, \sqrt{\%32}, \sqrt{\%33}]$) имеют одинаковую – единичную - длину, но принадлежат пространствам разной размерности. **Преобразование таких векторов равной длины, но разной размерности осуществляется операцией «тензор-унитарного преобразования», введенной докладчиком** [doi: 10.20944/preprints202208.0435.v1]. Она напоминает известные операции унитарного (или ортогонального) преобразования, которые сохраняют длины векторов, преобразуемых в пространстве **фиксированной** размерности. Унитарные (или ортогональные) преобразования играют ключевую роль для квантового компьютинга, где все вычисления производятся на их основе, и любое унитарное преобразование может служить в качестве гейта (логического вентиля). Тензор-унитарные преобразования интересны для развития компьютинга.

Мы все привыкли - по литературным текстам и строчной развертке сигналов в телевидении – к передаче информации в виде **одиночной** строки. Результативность метода стохастической матрешки при анализе геномных ДНК показывает, что **передача генетической информации опирается на принципиально иной способ развертки информации, использующий многоярусный пакет строк, стохастически взаимно связанных**.

Представляемые исследования близки к цели проф. **В.В. Налимова**: **построить вероятностную модель языка, а затем и сознания в целом** (https://ru.wikipedia.org/wiki/Налимов,_Василий_Васильевич)

В.В. Налимов (1910-1997)

Бинарные оппозиции Ю.Б.Румера и матричная генетика

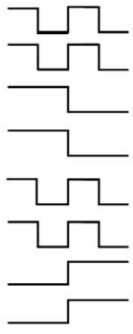
Ю.Б. Румер (1901-1985)

Известный физик-теоретик Ю.Б. Румер опубликовал в ДАН СССР в 1968 еще один вид бинарных оппозиций в системе генетического кодирования. Речь идет о делении множества 64 триплетов нуклеотидов по их кодовой нагрузке на два равных подмножества: 32 триплетов, начинающихся с 8 «сильных» дуплетов (CC, CT, CG, AC, TC, GC, GT, GG), и 32 триплетов, начинающихся с остальных 8 дуплетов («слабых»). Посмотрим, как в алфавитных матрицах ДНК располагаются полиплеты, начинающиеся с сильных дуплетов.

Обозначим сильные дуплеты и полиплеты с сильными корнями черным цветом со знаком «плюс», а остальные – белым со знаком «минус». Получаем мозаичные матрицы с симметриями, причем мозаики строк являются меандрами и представляются меандровыми функциями Радемахера, известными в теории вероятностей и цифровой информатике.

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG



	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	CCCC	CCCA	CCAC	CCAA	CACC	CACA	CAAC	CAAA	ACCC	ACCA	ACAC	ACAA	AACC	AACA	AAAC	AAAA
0001	CCCT	CCCG	CCAT	CCAG	CACT	CACG	CAAT	CAAG	ACCT	ACCG	ACAT	ACAG	AACT	AACG	AAAT	AAAG
0010	CCTC	CCTA	CCGC	CCGA	CATC	CATA	CAGC	CAGA	ACTC	ACTA	ACGC	ACGA	AATC	AATA	AAGC	AAGA
0011	CCTT	CCTG	CCGT	CCGG	CATT	CATG	CAGT	CAGG	ACTT	ACTG	ACGT	ACGG	AATT	AATG	AAGT	AAGG
0100	CTCC	CTCA	CTAC	CTAA	CGCC	CGCA	CGAC	CGAA	ATCC	ATCA	ATAC	ATAA	AGCC	AGCA	AGAC	AGAA
0101	CTCT	CTCG	CTAT	CTAG	CGCT	CGCG	CGAT	CGAG	ATCT	ATCG	ATAT	ATAG	AGCT	AGCG	AGAT	AGAG
0110	CTTC	CTTA	CTGC	CTGA	CGTC	CGTA	CGGC	CGGA	ATTC	ATTA	ATGC	ATGA	AGTC	AGTA	AGGC	AGGA
0111	CTTT	CTTG	CTGT	CTGG	CGTT	CGTG	CGGT	CGGG	ATTT	ATTG	ATGT	ATGG	AGTT	AGTG	AGGT	AGGG
1000	TCCC	TCCA	TCAC	TCAA	TACC	TACA	TAAC	TAAA	GCCC	GCCA	GCAC	GCAA	GACC	GACA	GAAC	GAAA
1001	TCCT	TCCG	TCAT	TCAG	TACT	TACG	TAAT	TAAG	GCCT	GCCG	GCAT	GCAG	GACT	GACG	GAAT	GAAG
1010	TCTC	TCTA	TCGC	TCGA	TATC	TATA	TAGC	TAGA	GCTC	GCTA	GCGC	GCGA	GATC	GATA	GAGC	GAGA
1011	TCTT	TCTG	TCGT	TCGG	TATT	TATG	TAGT	TAGG	GCTT	GCTG	GCGT	GCGG	GATT	GATG	GAGT	GAGG
1100	TTCC	TTCA	TTAC	TTAA	TGCC	TGCA	TGAC	TGAA	GTCC	GTCA	GTAC	GTAA	GGCC	GGCA	GGAC	GGAA
1101	TTCT	TTCG	TTAT										GGCT	GGCG	GGAT	GGAG
1110	TTTC	TTTA	TTGC										GGTC	GGTA	GGGC	GGGA
1111	TTTT	TTTG	TTGT										GGTT	GGTG	GGGT	GGGG

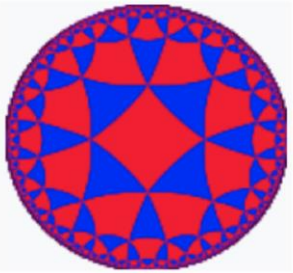
28 / 520

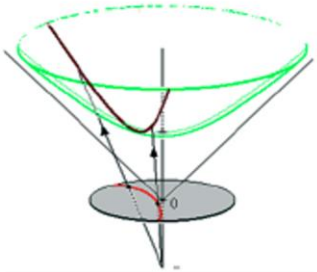
Соответствующие мозаичные матрицы процентов n-плетов водородных связей оказываются представлениями 2^n-мерных чисел алгебры сплит-кватернионов Кокла и ее расширений. Например, матрица процентов 4 дуплетов является суммой 4 разреженных матриц, набор которых замкнут относительно умножения и определяет таблицу умножения алгебры 4-мерных сплит-кватернионов Кокла (https://en.wikipedia.org/wiki/Split-quaternion).

Сплит-кватернионы давно используются в дисковой модели Пуанкаре для описания гиперболических движений в гиперболической геометрии Лобачевского (показан символ модели Пуанкаре).

$$\begin{vmatrix} \%33, & -\%32, & \%23, & -\%22 \\ \%32, & \%33, & -\%22, & -\%23 \\ \%23, & -\%22, & \%33, & -\%32 \\ -\%22, & -\%23, & \%32, & \%33 \end{vmatrix} = \%33 \begin{vmatrix} 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \\ 0,0,0,1 \end{vmatrix} + \%32 \begin{vmatrix} 0,-1,0,0 \\ 1,0,0,0 \\ 0,0,0,-1 \\ 0,0,1,0 \end{vmatrix} + \%22 \begin{vmatrix} 0,0,0,-1 \\ 0,0,-1,0 \\ 0,-1,0,0 \\ -1,0,0,0 \end{vmatrix} + \%23 \begin{vmatrix} 0,0,1,0 \\ 0,0,0,-1 \\ 1,0,0,0 \\ 0,-1,0,0 \end{vmatrix}$$

*	e_0	e_1	e_2	e_3
e_0	e_0	e_1	e_2	e_3
e_1	e_1	$-e_0$	e_3	$-e_2$
e_2	e_2	$-e_3$	e_0	$-e_1$
e_3	e_3	e_2	e_1	e_0





Быстрое преобразование Фурье и алгебраическая бит-реверсивная голография в цифровой информатике используют прочтение бинарных номеров в обратном порядке. Аналогичная реверсия бинарных номеров столбцов и строк данных мозаичных матриц приводит к алфавитным матрицам неожиданного блочного типа:

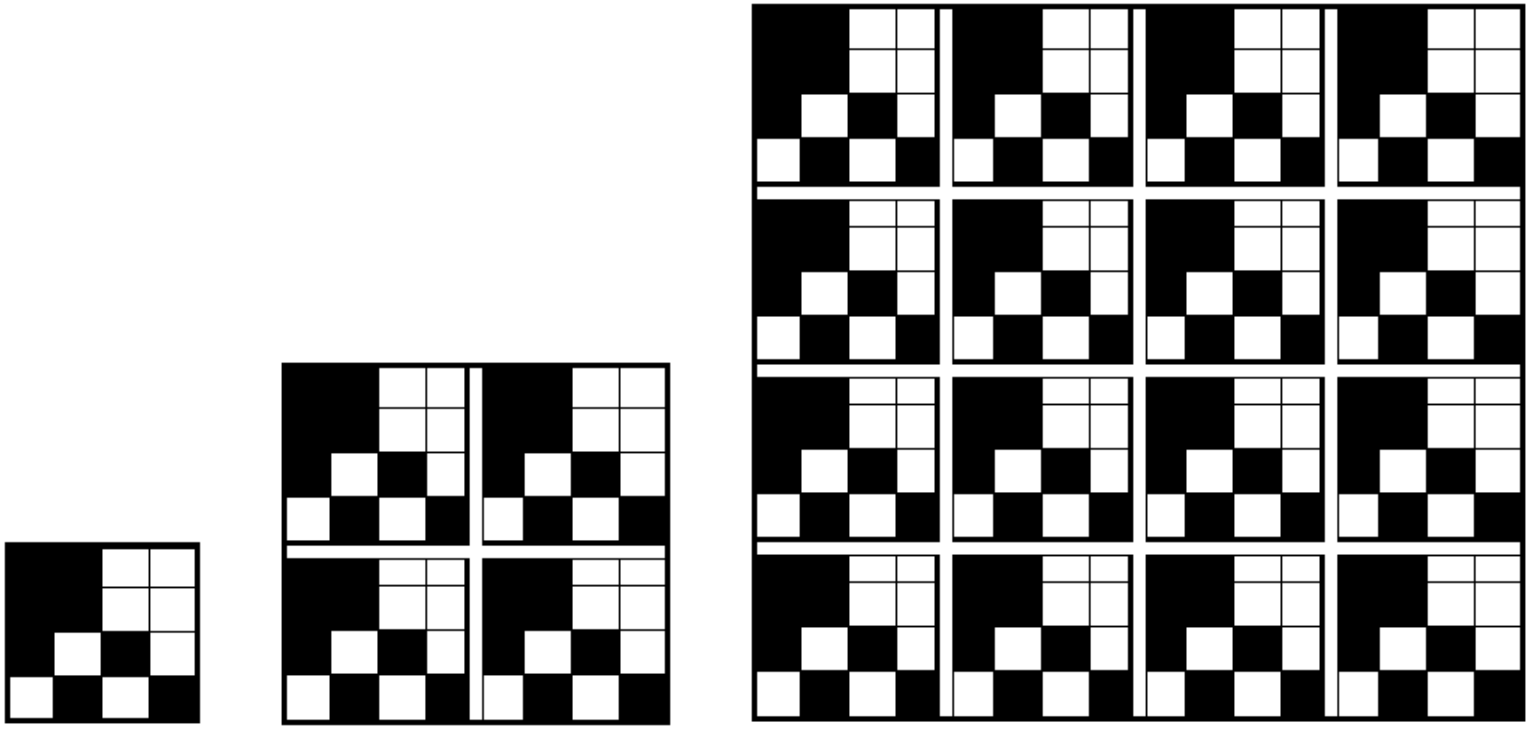
	00	10	01	11
00	CC	AC	CA	AA
10	TC	GC	TA	GA
01	CT	AT	CG	AG
11	TT	GT	TG	GG

	000	100	010	110	001	101	011	111
000	CCC	ACC	CAC	AAC	CCA	ACA	CAA	AAA
100	GCC	TCC	GAC	TAC	GCA	TCA	GAA	TAA
010	CGC	AGC	CTC	ATC	CGA	AGA	CTA	ATA
110	GGC	TGC	GTC	TTC	GGA	TGA	GTA	TTA
001	CCG	ACG	CAG	AAG	CCT	ACT	CAT	AAT
101	GCG	TCG	GAG	TAG	GCT	TCT	GAT	TAT
011	CGG	AGG	CTG	ATG	CGT	AGT	CTT	ATT
111	GGG	TGG	GTG	TTG	GGT	TGT	GTT	TTT

	0000	1000	0100	1100	0010	1010	0110	1110	0001	1001	0101	1101	0011	1011	0111	1111
0000	CCCC	ACCC	CACC	AACC	CCAC	ACAC	CAAC	AAAC	CCCA	ACCA	CACA	AACA	CCAA	ACAA	CAAA	AAAA
1000	TCCC	GCCC	TACC	GACC	TCAC	GCAC	TAAC	GAAC	TCCA	GCCA	TACA	GACA	TCAA	GCAA	TAAA	GAAA
0100	CTCC	ATCC	CGCC	AGCC	CTAC	ATAC	CGAC	AGAC	CTCA	ATCA	CGCA	AGCA	CTAA	ATAA	CGAA	AGAA
1100	TTCC	GTCC	TGCC	GGCC	TTAC	GTAC	TGAC	GGAC	TTCA	GTCA	TGCA	GGCA	TTAA	GTAA	TGAA	GGAA
0010	CCTC	ACTC	CATC	AATC	CCGC	ACGC	CAGC	AAGC	CCTA	ACTA	CATA	AATA	CCGA	ACGA	CAGA	AAGA
1010	TCTC	GCTC	TATC	GATC	TCGC	GCGC	TAGC	GAGC	TCTA	GCTA	TATA	GATA	TCGA	GCGA	TAGA	GAGA
0110	CTTC	ATTC	CGTC	AGTC	CTGC	ATGC	CGGC	AGGC	CTTA	ATTA	CGTA	AGTA	CTGA	ATGA	CGGA	AGGA
1110	TTTC	GTTC	TGTC	GGTC	TTGC	GTGC	TGGC	GGGC	TTTA	GTTA	TGTA	GGTA	TTGA	GTGA	TGGA	GGGA
0001	CCCT	ACCTT	CACT	AACT	CCAT	ACAT	CAAT	AAAT	CCCG	ACCG	CACG	AACG	CCAG	ACAG	CAAG	AAAG
1001	TCCT	GCCT	TACT	GACT	TCAT	GCAT	TAAT	GAAT	TCCG	GCCG	TACG	GACG	TCAG	GCAG	TAAG	GAAG
0101	CTCT	ATCT	CGCT	AGCT	CTAT	ATAT	CGAT	AGAT	CTCG	ATCG	CGCG	AGCG	CTAG	ATAG	CGAG	AGAG
1101	TTCT	GTCT	TGCT	GGCT	TTAT	GTAT	TGAT	GGAT	TTCG	GTCG	TGCG	GGCG	TTAG	GTAG	TGAG	GGAG
0011	CCTT	ACTT	CATT	AATT	CCGT	ACGT	CAGT	AAGT	CCTG	ACTG	CATG	AATG	CCGG	ACGG	CAGG	AAGG
1011	TCTT	GCTT	TATT	GATT	TCGT	GCGT	TAGT	GAGT	TCTG	GCTG	TATG	GATG	TCGG	GCGG	TAGG	GAGG
0111	CTTT	ATTT	CGTT	AGTT	CTGT	ATGT	CGGT	AGGT	CTTG	ATTG	CGTG	AGTG	CTGG	ATGG	CGGG	AGGG
1111	TTTT	GTTT	TGTT	GGTT	TTGT	GTGT	TGGT	GGGT	TTTG	GTTG	TGTG	GGTG	TTGG	GTGG	TGGG	GGGG

Мозаики всех этих возникающих алфавитных матриц вероятностей *n*-плетов являются повторением мозаики матрицы 4 дуплетов, образуя «матричные кристаллы».

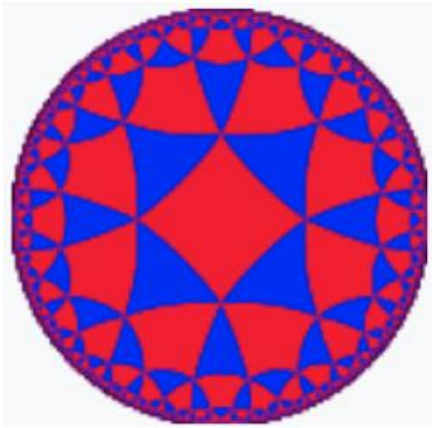
Этот «универсальный» повторяющийся блок, показанный внизу слева, оказывается новым матричным представлением тех же сплит-кватернионов Кокла, как показывает следующий слайд.

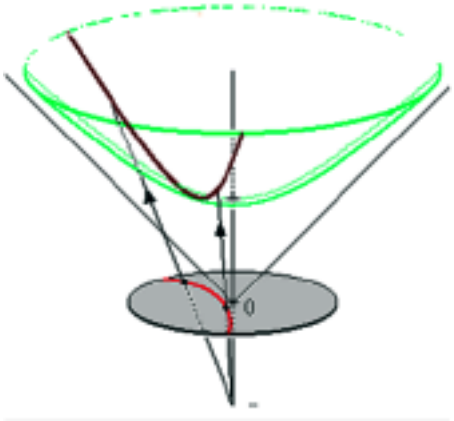


1	1	-1	-1
1	1	-1	-1
1	-1	1	-1
-1	1	-1	1

$$= \begin{vmatrix} 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \\ 0,0,0,1 \end{vmatrix} + \begin{vmatrix} 0,0,-1,0 \\ 0,0,0,-1 \\ 1,0,0,0 \\ 0,1,0,0 \end{vmatrix} + \begin{vmatrix} 0,0,0,-1 \\ 0,0,-1,0 \\ 0,-1,0,0 \\ -1,0,0,0 \end{vmatrix} + \begin{vmatrix} 0,1,0,0 \\ 1,0,0,0 \\ 0,0,0,-1 \\ 0,0,-1,0 \end{vmatrix} = j_0+j_1+j_2+j_3$$

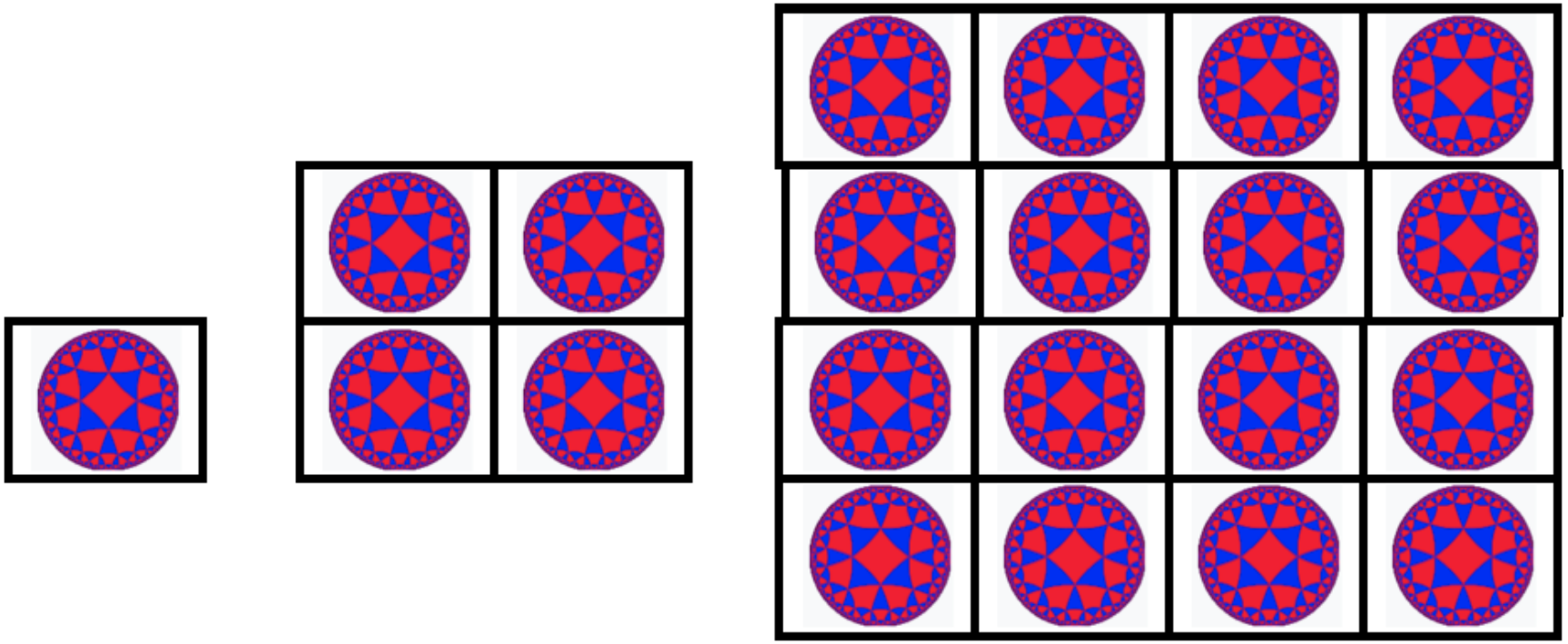
*	j_0	j_1	j_2	j_3
j_0	j_0	j_1	j_2	j_3
j_1	j_1	$-j_0$	j_3	$-j_2$
j_2	j_2	$-j_3$	j_0	$-j_1$
j_3	j_3	j_2	j_1	j_0



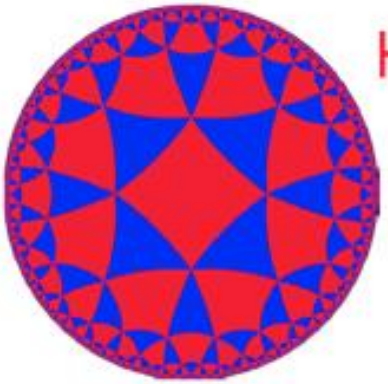


Этот блок является суммой четырех разреженных матриц, набор которых замкнут относительно умножения и также соответствует таблице умножения алгебры 4-мерных сплит-кватернионов Кокла (1849г.).

Используя символ дисковой модели Пуанкаре, получаем артистическое представление возникших «матричных кристаллов»:



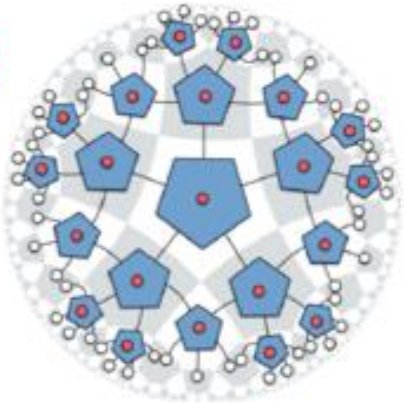
Это семейство блочно-унифицированных матриц ДНК-алфавитов неожиданно перекликается с темой «**голографических квантовых кодов, корректирующих ошибки**». Эта тема развивается в Калифорнийском Технологическом Институте США в связи с той же дисковой моделью Пуанкаре и ее замощениями:



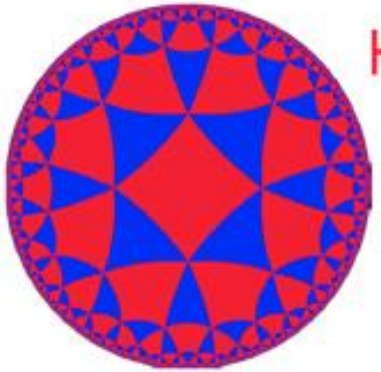
Holographic correspondence

Quantum error correction

Are they closely related?



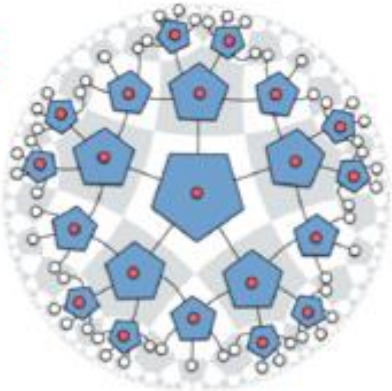
(Картинка из презентации на Американском Физическом Обществе в 2016г.http://theory.caltech.edu/~preskill/talks/APS-March-2016-preskill.pdf).



Holographic correspondence

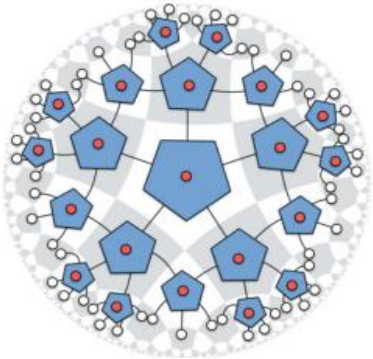
Quantum error correction

Are they closely related?

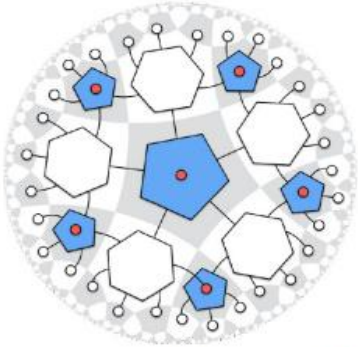


Разработка этой голографической темы возглавляется J.Preskill, директором Института Квантовой Информатики и Материи в Калтехе. Данная тема включает также рассмотрение пространства-времени как квантового помехоустойчивого кода: «Is spacetime a quantum error-correcting code?» (arXiv:1503.06237).

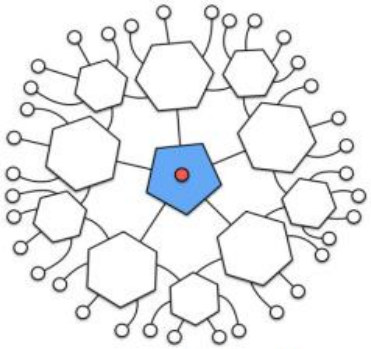
Holographic quantum codes



pentagon code



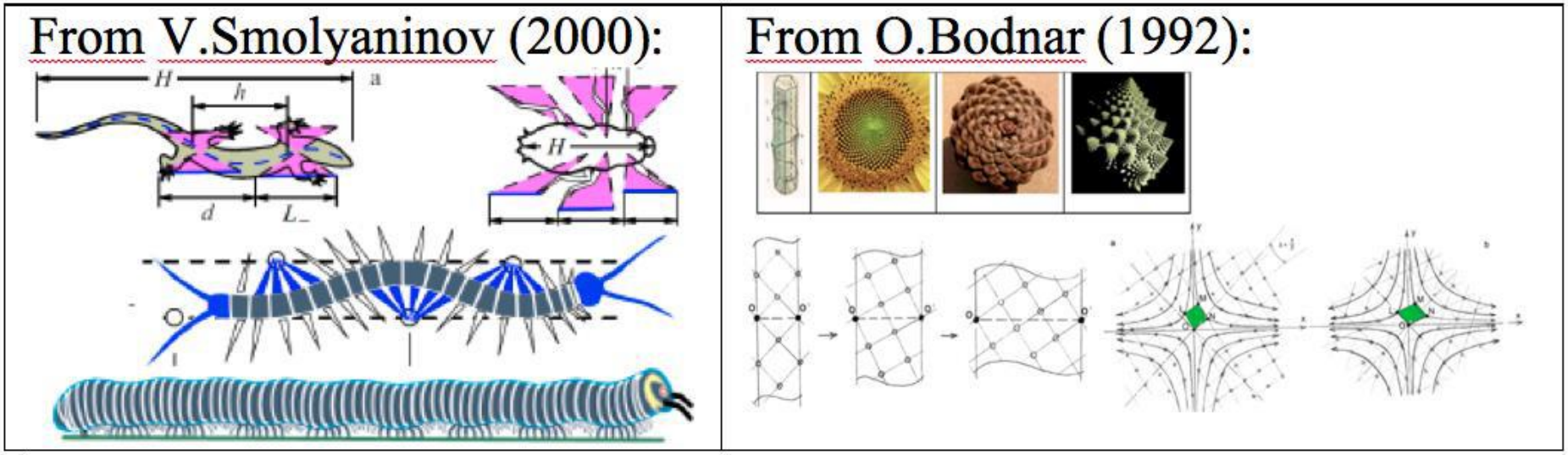
pentagon/hexagon code



one encoded qubit

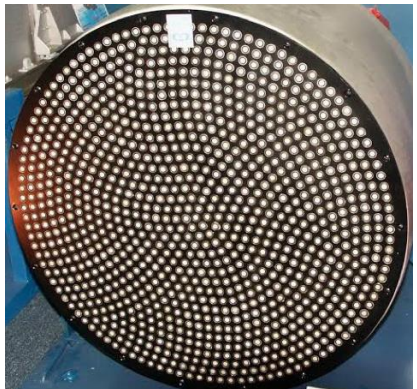
Выявление алгебраической связи системы генетического кодирования с гиперболической геометрией позволяет видеть генетическую основу в известных сопряжениях с этой геометрией следующих физиологических явлений:

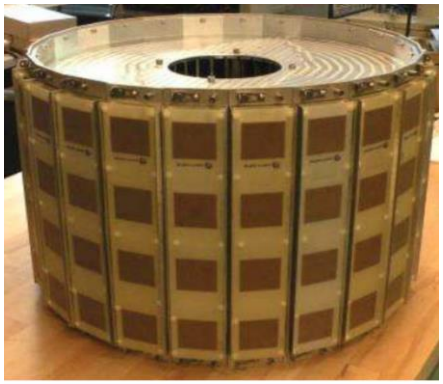
- гиперболической геометрии пространства зрительного восприятия [Luneburg, 1950; Kienle, 1964];
- пространственно-временной организации локомоций человека и животных [Смолянинов, УФН, 2000];
- морфогенетических феноменах филлотаксиса [Боднар, 1992, 1994]. Эти авторы не предполагали связи открываемых ими биофеноменов со структурами молекулярного генетического кодирования.



ДОКТРИНА ЭНЕРГО-ИНФОРМАЦИОННОЙ ЭВОЛЮЦИИ НА ОСНОВЕ БИО-АНТЕННЫХ РЕШЕТОК.

Алгебраические связи между матрицами вероятностей *n*-плетов в геномных ДНК оказались аналогичными формализмам известной тензорно-матричной теории цифровых антенных решеток, называемых также Умными Антеннами [Слюсарь В.И., 1999 и далее]. Антенные решетки состоят из множества согласованных антенн, каждая из которых поглощает или излучает волны определенного диапазона. Антенные решетки кардинально превосходят возможности отдельных антенн по приему и передаче информации, а потому нашли тысячи применений в технике (рисунки взяты из Интернета).

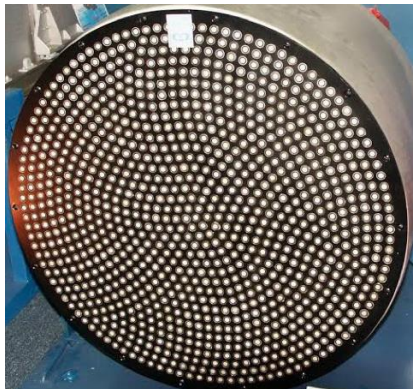


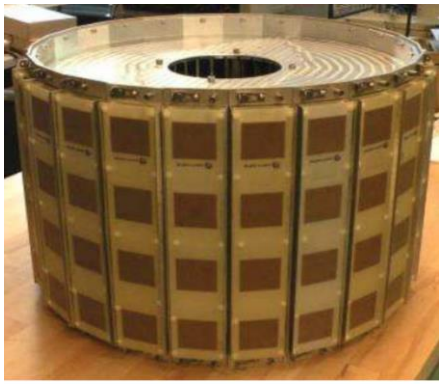


Пример UAR Alcatel-Lucent, 2011 год

Современная наука связывает с антенными решетками революцию в компьютерах (биофотоника) и энергетике (солнечная энергетика на наноантенных решетках).

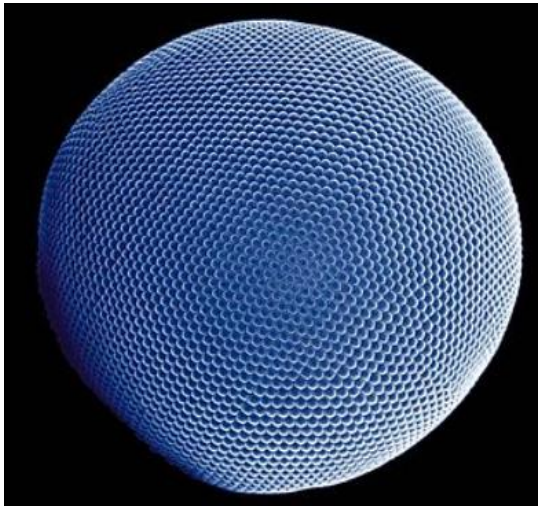
Но в биологии не было известно ни одной публикации о роли био-антенных решеток в живых телах. Названная алгебраическая аналогия позволила докладчику выдвинуть и обосновать доктрину энерго-информационной эволюции на основе био-антенных решеток и их волновой деятельности в организмах [Petoukhov, Biosystems, 2022]. Эта доктрина опирается на множество выявленных фактов использования аналогов антенных решеток и их волновых функций в организмах, многие ткани которых являются пьезоэлектриками. Приведем примеры.

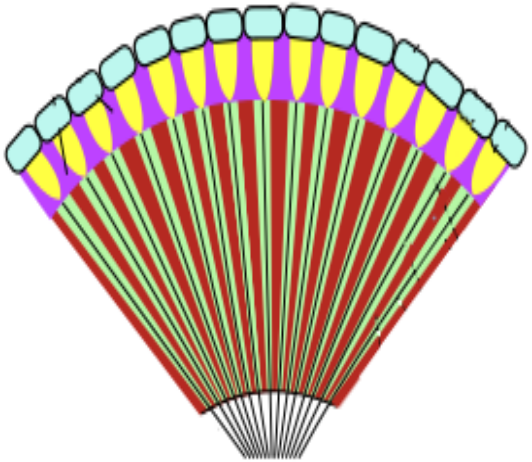




Пример UAP Alcatel-Lucent, 2011 год

ФАСЕТОЧНЫЕ ГЛАЗА

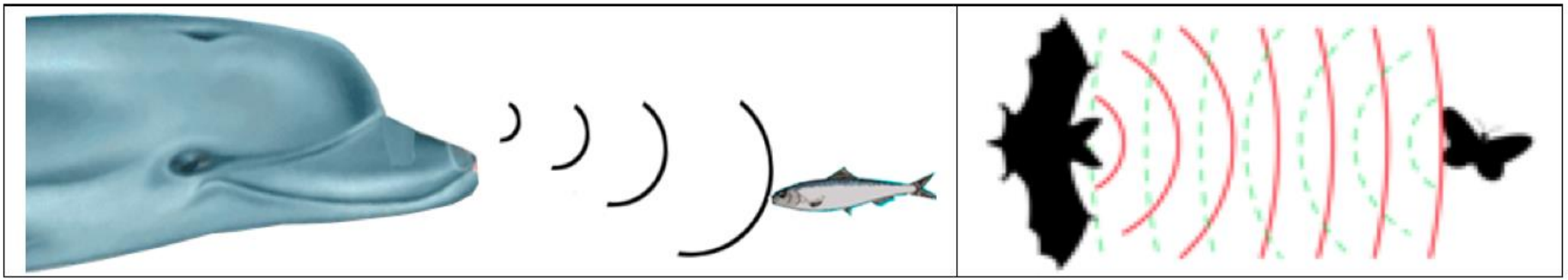




Насекомые и многие беспозвоночные получают зрительную информацию через фасеточные глаза как био-антенные решетки приема электромагнитных волн.

(Все изображения взяты из Интернета).

ВРОЖДЕННЫЕ ФЕНОМЕНЫ ЭХОЛОКАЦИИ



Многие организмы обладают врожденной способностью к эхолокации. Например, эхолокация позволяет дельфинам и летучим мышам определять дистанцию, размеры и форму объектов по аналогии с Умными Антеннами (Smart Antennas) в технике. Тело дельфина имеет множество гидроакустических рецепторов, образующее широкополосную приемную антенную решетку.

Врожденные феномены электролокации

Многие организмы наделены способностью электролокации для поиска, анализа и коммуникаций. Соответствующие органы генерации и приема электромагнитных сигналов распределены вдоль их тела, образуя аналоги антенных решеток.

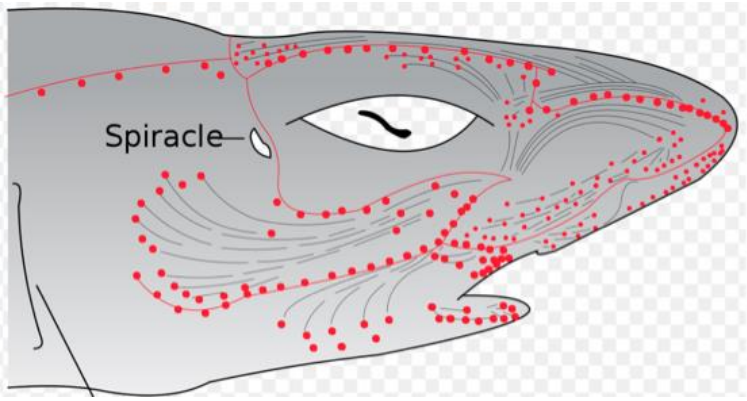


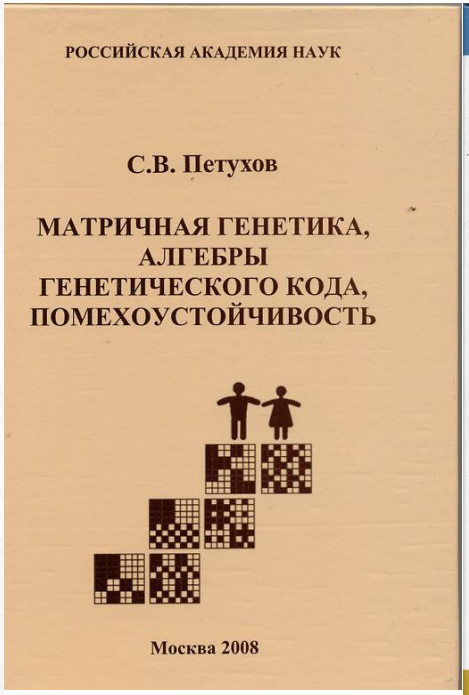
Рисунок показывает распределение рецепторов электромагнитного поля на голове акулы (взято из Интернета).

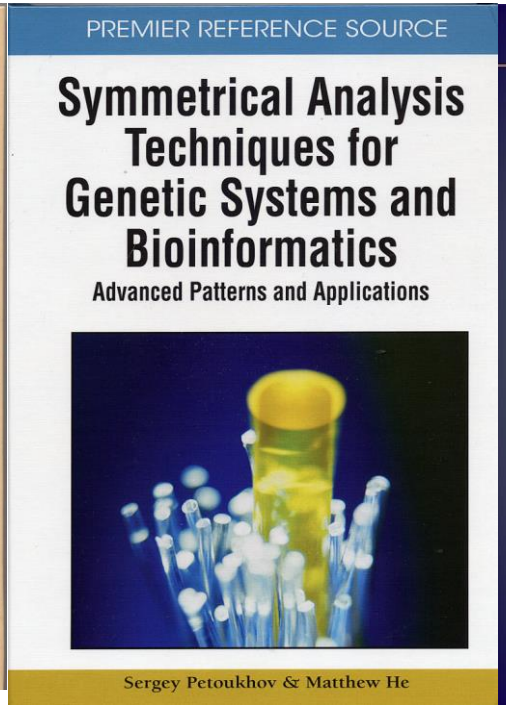
Врожденные наборы биофотонных кристаллов

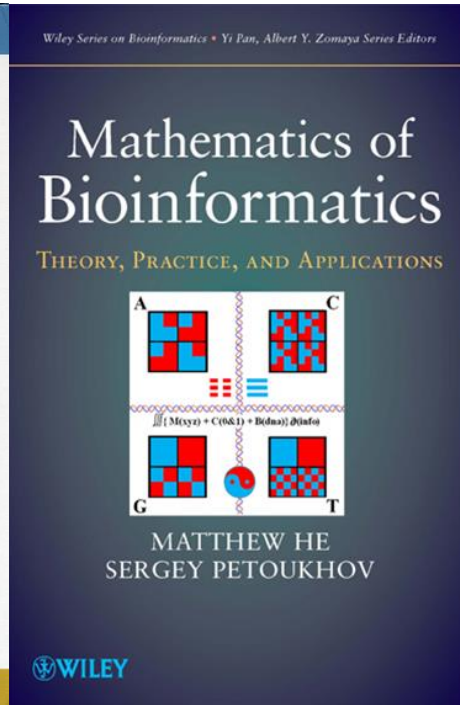
Фотонные кристаллы формируют видовой узор на крыльях бабочек и перьях павлинов. Современная техника широко использует фотонные кристаллы для манипуляций с пучками фотонов и создания нано-антенных решеток.

С био-антенными решетками в живом связаны интеллектуальная и вибро-акустическая деятельность, помехоустойчивость, извлечение слабых сигналов на фоне шума, эффекты поляризации, направленные потоки энергии, самоорганизация и пр. [Petoukhov, Biosystems, 2022].

Дополнительные материалы к докладу имеются на сайте http://petoukhov.com/ .







ЗАКЛЮЧЕНИЕ

1) Универсальные структуры молекулярно-генетической информатики и наследуемых физиологических систем образуют обширное поле для плодотворного применения алгебро-геометрических методов в интересах многих наук и дальнейшего развития алгебраической биологии, теории систем, а также искусственного интеллекта на базе патентов живой природы;

2) Результаты исследований по матричной генетике желательно учитывать при создании разнообразных систем искусственного интеллекта с перспективой развития геноморфного искусственного интеллекта, максимально приближенного к естественному генетическому интеллекту, элементы которого наследуются из поколения в поколение.