

COMPLETE QUANTUM INFORMATION IN THE DNA GENETIC CODE

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡],
MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

ABSTRACT. We find that the degeneracies and many peculiarities of the DNA genetic code may be described thanks to two closely related (fivefold symmetric) finite groups. The first group has signature $G = \mathbb{Z}_5 \rtimes H$ where $H = \mathbb{Z}_2.S_4 \cong 2O$ is isomorphic to the binary octahedral group $2O$ and S_4 is the symmetric group on four letters/bases. The second group has signature $G = \mathbb{Z}_5 \rtimes GL(2, 3)$ and points out a threefold symmetry of base pairings. For those groups, the representations for the 22 conjugacy classes of G are in one-to-one correspondence with the multiplets encoding the proteinogenic amino acids. Additionally, most of the 22 characters of G attached to those representations are informationally complete. The biological meaning of these coincidences is discussed.

Arxiv: q-bio.OT, quant-ph, math.GR, math.AG

PACS: 02. 20.-a, 82.39.Pj, 87.10.Vg, 02.10.De

MSC codes: 20C15, 92D20, 20E45, 14H52, 14G05

Keywords: DNA genetic code, informationally complete characters, finite groups, hyperelliptic curve

1. INTRODUCTION

La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i qualimezi impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto. (Galileo Galilei (1564-1642), Il Saggiatore, cap. 6).

Until now, deciphering the code of life [1]-[8] –the genetic code– was still unsuccessful although mathematical theories have been proposed before [9]-[13]. How is our new attempt different from earlier trials? Our mathematical approach lies at the crossroads of finite group theory and quantum information in the line of other papers mainly devoted to quantum computing [14] but also focused on elementary particles [15].

Life cells need a macromolecule called deoxyribonucleic (or DNA) packed in a chromosome during the cell mitosis. But DNA unwinds to be copied during DNA replication or when its code is used to make proteins. DNA is an helix consisting of two parallel polynucleotide chains carrying genetic instructions in 4 nitrogenous bases for the growth and reproduction of all living organisms. The genetic code is organized in triples of bases called

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡], MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

codons. There are $4^3 = 64$ codons but only 20 standard amino acids, meaning a high redundancy of the code. In our approach, we use the characters and the corresponding representations of a well defined finite group G to explain the mapping of the nitrogenous bases to amino acids. The group characters may be used to build minimal and informationally complete quantum measurements, one for each character and the corresponding amino acid.

In Section 2, one details the discovery and main properties of DNA and the genetic code. In Section 3, one recalls how minimal informationally complete quantum measurements are performed and the necessary elements of character theory of finite groups. In Section 4, the assignment of characters of the small group $G = (240, 105)$ to amino acids is detailed. In Section 5, we provide a justification of the fact that minor and major grooves in the DNA double helix have periods whose ratio is the Golden ratio. This is based on the study of points over a hyperelliptic curve occurring in the character table.

2. DNA AND THE GENETIC CODE

The discovery of DNA is attributed to Watson and Crick in 1953 [1]. But the phosphorous-containing substance now called DNA was first isolated by F. Miescher in 1869 in the nuclei of white blood cells under the name of ‘nuclein’ (a nucleic acid) paving the way of its recognition as the carrier of inheritance [2]. Only in 1909, P. Levene found that DNA contains pentoses such adenine (A), guanine (G), thymine (T), cytosine (C), a deoxyribose and a phosphate group. At that time, it was still believed that the protein component of chromosomes was the true basis of heredity. The recognition that DNA rather than proteins could be the genetic material was suggested by the E. Chargaff’s rules, proposed in 1940 that the four bases are present in the percentages: $A \approx T \approx 30\%$ and $G \approx C \approx 20\%$ for all species [3]. Subsequent X-ray crystallography work by English researchers R. Franklin and M. Wilkins contributed to Watson and Crick’s derivation of the three-dimensional, double-helical model for the structure of DNA [4].

As shown in Fig. 1, DNA is a double helix, with the two strands connected by hydrogen bonds. In the most current form of DNA (called B-DNA), the ratio of the diameter $D \approx 21\text{\AA}$ to the period $l + L$ and the ratio between the minor groove $l \approx 13\text{\AA}$ and the major groove $L \approx D \approx 21\text{\AA}$ are both close to the golden ratio $\phi = (\sqrt{5} - 1)/2 \approx 0.618$ [17].

G. Gamow (also a cosmologist) observed that the $4^3 = 64$ possible permutations of the four DNA bases A, T, G and C, taken three at a time (as codons), could be reduced to 20 distinct combinations and might code for the twenty amino acids which, he suggested, might well be the sole constituents of all proteins [5]. But the lack of overlapping of codons (not assumed by Gamow) and the demonstration that the genetic code is made up of a series of three base pair codons, which code for individual amino acids, dates back to 1961 with the Crick, Brenner et al. experiment [6]. Now we know that the peculiar non ambiguous and non overlapping assignment of all 64 codons to the 20 amino acids is nearly universal [7].

The ‘genetic code’ is the set of rules used by living cells to translate information encoded within genetic material (DNA or messenger RNA sequences

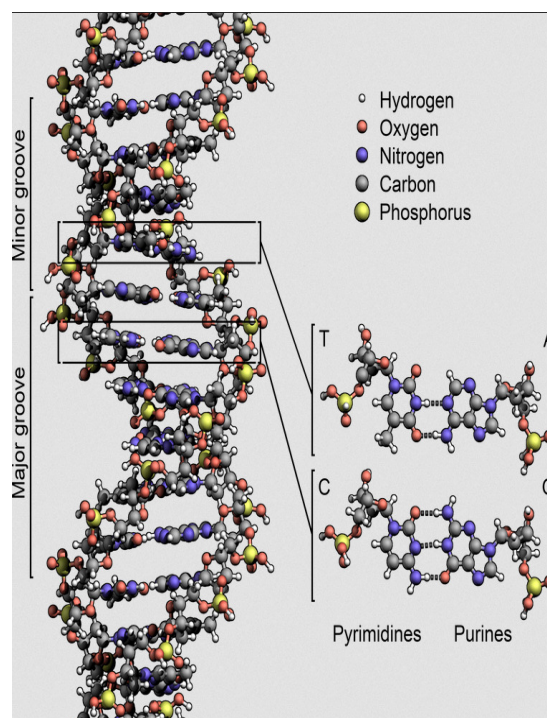


FIGURE 1. The double-helical structure of DNA and base pairing. Complementary bases are held together as a pair by hydrogen bonds. Two hydrogen bonds connect T to A; three hydrogen bonds connect G to C. The image is borrowed to Wikipedia [16].

of nucleotide triplets, or codons) into proteins. Translation is accomplished by the ribosome, which links amino acids in an order specified by messenger RNA (mRNA), using transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA, three nucleotides at a time.

The codons which code for the same amino acids form multiplets and are organized as

- * Met, Trp: 2 singlets
- * Asn, Asp, Cys, Gln, Glu, His, Lys, Phe, Tyr: 9 doublets
- * Ile, Term: 2 triplets
- * Ala, Gly, Pro, Thr, Val: 5 quadruplets
- * Arg, Leu, Ser: 3 sextets.

A 21st proteinogenic amino acid Selenocysteine (symbol Sec) was found present in some enzymes and a 22nd amino acid Pyrrolysine (Symbol Pyl) in some archaea and bacteria.

The assignment of three-letter codons to the standard 20 amino acids is shown in Fig. 2a. It is clearly seen in this table that the assignment of codons to amino acids is not random but follows underlying rules to be discovered.

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡], MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

Woese and collaborators introduced the concept of polar requirement to express stereochemical associations between amino acids and nucleobases in solution [8]. Taking dimethylpyridine (or DMP) as a solvent, the polar requirement is found as in Fig. 2b [18, 19]. For more recent predictors of the stereochemical associators the reader may read reference [20].

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

(a)

		Middle Position			
		U	C	G	A
First Position	U	Phe 5.0	Ser 7.5	Cys 4.8	Tyr 5.4
	C	Leu 4.9	Pro 6.6	Arg 9.1	His 8.4 Gln 8.6
	G	Val 5.6	Ala 7.0	Gly 7.9	Asp 13.0 Glu 12.5
	A	Ile 4.9 Met 5.3	Thr 6.6	Ser 7.5 Arg 9.1	Asn 10.0 Lys 10.1

(b)

FIGURE 2. (a) the generic codon table for RNA, (b) the codon table colored by amino acid polar requirement values [19].

3. SYMMETRIES AND QUANTUM INFORMATION FROM THE CHARACTERS OF A FINITE GROUP

There are two important ingredients in our approach. One topic follows from our earlier work on ‘magic state’ quantum computing [14] and consists of performing a generalized quantum measurement while keeping a complete information in the process, as explained in Sec. 3.1. The second (more recent) topic consists of finding the ‘magic states’ in the bank of characters of a selected finite group. The latter approach recalled in Section 3.2 was found successful in the context of symmetries of elementary matter particles [15].

3.1. Minimal informationally complete quantum measurements. Let \mathcal{H}_d be a d -dimensional complex Hilbert space and $\{E_1, \dots, E_m\}$ be a collection of positive semi-definite operators (POVM) that sum to the identity. Taking the unknown quantum state as a rank 1 projector $\rho = |\psi\rangle\langle\psi|$ (with $\rho^2 = \rho$ and $\text{tr}(\rho) = 1$), the i -th outcome is obtained with a probability given by the Born rule $p(i) = \text{tr}(\rho E_i)$. A minimal and informationally complete POVM (or MIC) requires d^2 one-dimensional projectors $\Pi_i = |\psi_i\rangle\langle\psi_i|$, with

$\Pi_i = dE_i$, such that the rank of the Gram matrix with elements $\text{tr}(\Pi_i\Pi_j)$, is precisely d^2 .

With a MIC, the complete recovery of a state ρ is possible at a minimal cost from the probabilities $p(i)$. In the best case, the MIC is symmetric and called a SIC with a further relation $|\langle\psi_i|\psi_j\rangle|^2 = \text{tr}(\Pi_i\Pi_j) = \frac{d\delta_{ij}+1}{d+1}$ so that the density matrix ρ can be made explicit [21].

In our earlier references, starting with [14], a large collection of MICs are derived. They correspond to Hermitian angles $|\langle\psi_i|\psi_j\rangle|_{i\neq j} \in A = \{a_1, \dots, a_l\}$ belonging to a discrete set of values of small cardinality l . They arise from the action of a Pauli group \mathcal{P}_d [22] on an appropriate magic state pertaining to the coset structure of subgroups of index d of a free group with relations.

In reference [15], an entirely new class of MICs in the Hilbert space \mathcal{H}_d , relevant for the lepton and quark mixing patterns, is obtained by taking fiducial/magic states as characters of a finite group G possessing d conjugacy classes and using the action of a Pauli group \mathcal{P}_d on them. The same approach is used in this paper.

3.2. The character table of a finite group and quantum information. Let G be a finite group, V a finite-dimensional vector space over a field F and let $r : G \rightarrow GL(V)$ be a representation of G on V . The character $\kappa_r(g)$ associates to an element $g \in G$ the trace of the corresponding matrix $r(g)$. A character κ_r is called irreducible if r is an irreducible representation. The degree of the character κ is called the dimension of the representation.

Assume that G is made of d conjugacy classes of elements. One introduces class functions on G from the ring of complex-valued functions on G that are constant on conjugacy classes. In fact we restrict ourselves to functions with values in cyclotomic fields [23].

The table of irreducible characters of G , or character table for short, is a way to summarize the properties of irreducible representations of elements g of G .

Let us take the case of the symmetric group S_4 also called the small group (24, 12) (number 12 in the range of 15 groups of order 24). The character table is in Table 1. It makes explicit the five irreducible characters of G , the corresponding dimension dim of the representation, as well as the size and order of an element in each class. For each class, we added the calculation of the rank of the Gram matrix that corresponds to the POVM obtained by the action of the 5-dit Pauli group over the irreducible character κ_i of the class. For the classes 3 to 5, the rank of the Gram matrix is $d^2 = 25$ so that the POVM is a MIC.

As a second example, one takes the binary octahedral group $2O$ also called the small group (48, 28) (number 28 in the range of 52 groups of order 48). Group $2O$ is an important ingredient of our model of the genetic code, as shown in section 4. The character table is shown in Table 2. It makes explicit the eight irreducible characters of G , the corresponding dimension of the representation, as well as the size and order of an element in each class. At the right hand side column, one finds the rank of the Gram matrix that corresponds to the POVM obtained by the action of the 8-dit Pauli

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡], MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

Class	1	2	3	4	5	Class
Size	1	3	6	8	6	Size
Order	1	2	2	3	4	Order
Char	dim					Gram
κ_1	1	1	1	1	1	5
κ_2	1	1	-1	1	-1	21
κ_3	2	2	0	-1	0	d^2
κ_4	3	-1	-1	0	1	d^2
κ_5	3	-1	1	0	-1	d^2

TABLE 1. The character table for the symmetric group $G = S_4$. There are five conjugacy classes of elements of dimensions 1, 1, 2, 3 and 3, respectively. The size and the order of an element in the class is as shown. The last five rows of the table are the irreducible characters κ_i , $i = 1..5$, of G . To each of them is associated the rank of the Gram matrix defined in subsection 3.1.

group over the irreducible character κ_i of the class. None of the characters of the group $2O$ is informationally complete.

Class	1	2	3	4	5	6	7	8	Class
Size	1	1	8	6	12	8	6	6	Size
Order	1	2	3	4	4	6	8	8	Order
Char	dim								Gram
κ_1	1	1	1	1	1	1	1	1	8
κ_2	1	1	1	1	-1	1	-1	-1	45
κ_3	2	2	-1	2	0	-1	0	0	56
κ_4	2	-2	-1	0	0	1	$\sqrt{2}$	$-\sqrt{2}$	59
κ_5	2	-2	-1	0	0	1	$-\sqrt{2}$	$\sqrt{2}$	59
κ_6	3	3	0	-1	1	0	-1	-1	59
κ_7	3	3	0	-1	-1	0	1	1	59
κ_8	4	-4	1	0	0	-1	0	0	59

TABLE 2. The character table for the symmetric group $G = 2O = (48, 28)$. There are eight conjugacy classes of elements of dimensions $1(\times 2)$, $2(\times 3)$, $3(\times 2)$ and 4. The size and the order of an element in the class is as shown. The last five rows of the table are the irreducible characters κ_i , $i = 1..8$, of G . To each of them is associated the rank of the Gram matrix defined in subsection 3.1.

4. SYMMETRIES AND QUANTUM INFORMATION IN THE GENETIC CODE

As shown in the previous section, there may exist several representations of a given dimension in a finite group. Does it exist a group G whose multiplets map to the multiplets of the genetic code in a satisfactorily way?

We shall impose the constraint that there are enough n -plets to embed the mappings to the proteogenomic amino acids for dimensions $n = 1, 2, 3$ and 4 and that there are at least 2 sextuplets. Small groups $(240, 105)$ and $(240, 106)$ (with 22 conjugacy classes) and $(240, 107)$ and $(240, 108)$ (with 28 conjugacy classes) are the group candidates with the smallest cardinality. All of them have two 6-dimensional representations. We could not find other candidate group of higher cardinality. The most economical groups are thus the former two groups and the latter two are different only in the sense that there are 6 one-dimensional representations instead of two.

The first group $(240, 105)$ has signature $G = \mathbb{Z}_5 \times H$ where $H = \mathbb{Z}_2.S_4 \cong 2O$ is isomorphic to the binary octahedral group $2O$ and S_4 is the symmetric group on four letters/bases. The second group $(240, 106)$ has signature $G = \mathbb{Z}_5 \times GL(2, 3)$ and points out a threefold symmetry of base pairings. For those groups, the representations for the 22 conjugacy classes of G are almost in one-to-one correspondence with the multiplets encoding the proteinogenic amino acids. The assignment is not perfect in the sense that there are only two representations of dimension 6 while there are three sextuplets in the genetic code.

$(240, 105)$ $\mathbb{Z}_5 \times (\mathbb{Z}_2.S_4)$ $\cong \mathbb{Z}_5 \times 2O$	dimension	1	1	2	2	2	2	2	2	2	2	2
	d-dit, d=22	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2
	amino acid	Met	Trp	Cys	Phe	Tyr	His	Gln	Asn	Lys	Glu	Asp
	order	1	2	3	4	4	5	5	6	8	8	10
	polar req.	5.3	5.2	4.8	5.0	5.4	8.4	8.6	10.0	10.1	12.5	13.0
$(240, 105)$	dimension	3	3	4	4	4	4	4	4	4	6	6
	d-dit, d=22	d^2	475	483	480	d^2	d^2	d^2	d^2	d^2	d^2	d^2
	amino acid	Ile	Stop	Leu, Pyl, Sec	Leu	Val	Pro	Thr	Ala	Gly	Ser	Arg
	order	10	15	15	15	15	20	20	30	30	30	30
	polar req.	4.9			4.9	5.6	6.6	6.6	7.0	7.9	7.5	9.1

TABLE 3. For the group $G = (240, 105) \cong \mathbb{Z}_5 \times 2O$, the table provides the dimension of the representation, the rank of the Gram matrix obtained under the action of the 22-dimensional Pauli group, the order of a group element in the class and a good assignment to an amino acid according to its polar requirement value. Bold characters are for faithful representations. There is an ‘exception’ for the assignment of the sextet ‘Leu’ that is assumed to occupy two 4-dimensional slots. All characters are informationally complete except for the ones assigned to ‘Stop’, ‘Leu’, ‘Pyl’ and ‘Sec’.

The 22 characters of the group $G = (240, 105)$ are investigated in Section 5. A sketch of the character properties of G is in Table 3. It includes the dimension of the representation associated to each character, the order of an element in each class, and additionally the rank of the Gram matrix when one builds a POVM based on the character as a magic/fiducial state as in section 3.1. Table 3 proposes a one-to-one assignment of the representations of G to amino acids. A given degeneracy of an amino acid in the genetic code corresponds to the dimension of the selected representations of G . The amino acids of a given degeneracy/dimension in the table are ordered according to their increasing polar requirement value (given in Fig. 2b) and

MICHEL PLANAT†, RAYMOND ASCHHEIM‡, MARCELO M. AMARAL‡, FANG FANG‡ AND KLEE IRWIN‡

(240,105) $\cong \mathbb{Z}_5 \times 2O$	dimension amino acid	1 Met	1 Trp	2 Cys	2 Phe	2 Tyr	2 His	2 Gln	2 Asn	2 Lys	2 Glue	2 Asp
	bases	AUG	UGG	UGU UGC	UUU UUC	UAU UAC	CAU CAC	CAA CAG	AAU AAC	AAA AAG	GAA GAG	GAU GAC
(240,105)	dimension amino acid	3 Ile	3 Stop	4 Leu Pyl,Sec	4 Leu	4 Val	4 Pro	4 Thr	4 Ala	4 Gly	6 Ser	6 Arg
	bases	AUU AUC AUA	UAA UAG UGA	UUA UUG UAG UGA	CUU CUC CUA CUG	GUU GUC GUA GUG	CCU CCC CCA CCG	ACU ACC ACA ACG	CGU GCC GCA GCG	GGU GGC GGA GGG	UCU UCC UCA UCG AGU AGC	CGU CGC GGA CGG AGA AGG

TABLE 4. Excerpt of Table 3 with the assignment of codons to amino acids according to the standard genetic code. There is no ambiguity in such assignments except for codons, UAG and UGA that code for ‘Stop’, as well as for ‘Pyl’ and ‘Sec’, respectively. According to our approach, this ambiguity is reflected in the characters being non informationally complete for the corresponding two slots.

(240,106) $\cong \mathbb{Z}_5 \times GL(2, 3)$	dimension	1	1	2	2	2	2	2	2	2	2	2
	order	1	2	2	3	4	5	5	6	8	8	10
	d-dit, d=22	d^2	d^2	483	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2
(240,106)	dimension	3	3	4	4	4	4	4	4	4	6	6
	order	10	15	15	15	15	20	20	30	30	30	30
	d-dit, d=22	482	d^2	d^2	d^2	d^2	d^2	d^2	d^2	d^2	440	440

TABLE 5. The group $(240, 106) \cong \mathbb{Z}_5 \times GL(2, 3)$ as another candidate for the assignments of amino acids. It has the same dimensions of representations and closely related group orders. One may postulate that the group $(240, 106)$ encodes D-amino acids while the group $(240, 105)$ encodes L-amino acids. For most naturally-occurring amino acids, the carbon alpha to the amino group has the L-configuration.

thus follows the order of a group element in the corresponding class. Most representations are informationally complete with the rank of the Gram matrix equal to d^2 . The exceptions are for the 3-dimensional slot for the Stop codon and the 4-dimensional slot that we associate to the amino-acid Leu and the 2 extra amino acids Pyl and Sec. There is an ambiguity in the assignment of codons UAG and UGA that code for Stop, and for Pyl (with the codon UAG) and Sec (with the codon UGA). Thus the slight ambiguity of nature in coding the amino acids is recovered in our approach. The assignments to codons is displayed in detail in Table 4. According to our approach, this ambiguity is reflected into three characters being non informationally complete.

Another possible assignment of amino acids to the representations of group $(240, 106)$ is possible. This group has similar dimensions of the representations than group $(240, 105)$ with slight differences in the order of an

element in each class and the rank of the Gram matrix for the corresponding POVM, as shown in Table 5. One may postulate that it encodes the D-amino acids while the group $(240, 105)$ encodes proteinogenic L-amino acids. For most naturally-occurring amino acids, the carbon alpha to the amino group has the L-configuration. D-amino acids are most occasionally found in nature as residues in proteins.

Normal subgroups $(48, 28) \cong 2O$ of the group $(240, 105)$ and $GL(2, 3) \cong (48, 29)$ of the group $(240, 106)$ may be used to encode most of the so-called essential amino-acids as shown in Table 6. One observes that apart from two exceptions for the 4-dimensional slots of $GL(2, 3)$ the characters of these groups are not informationally complete.

	dimension order	1	1	2	2	2	3	3	4
$(48,28)=2O$	d-dit, d=8	1	2	3	4	4	6	8	8
$(48,29)=GL(2,3)$	d-dit, d=8	8	45	56	59	59	59	59	59
	amino acid polar req.	Met	Trp	Phe	His	Lys	Ile	Stop	Thr
		5.3	5.2	5.0	8.4	10.1	4.9		6.6

TABLE 6. The assignment of representations of groups $2O$ or $GL(2, 3)$ to essential amino acids. There are two extra essential amino acids Val and Leu but they are not strictly essential from a metabolic perspective. Strictly indispensable amino acids are Trp, Lys and Thr which cannot be submitted to transamination.

5. SYMMETRIES OF THE GENETIC CODE, THE GOLDEN RATIO AND A HYPERELLIPTIC CURVE

If one accepts the validity of the aforementioned group theoretical model of the genetic code, one can provide clues why nature selected these particular symmetries. One clue is contained in the detailed structure of the characters of the group $(240, 105)$ and its cousin group $(240, 106)$. The golden ratio is playing a role in accordance with the DNA structure.

The list of 22 characters for the group $(240, 105)$ is in Table 7. A closely related table (not shown) is found for the group $(240, 106)$. The entries in the character table are shown to be either rational numbers, quadratic irrationalities such as $z_1 = -(\sqrt{5} + 1)/2$, $z_2 = \sqrt{5} - 1$, $z_3 = 3(1 + \sqrt{5})/2$, $z_4 = \sqrt{2}$ and their algebraic conjugates, or $\pm r_i$ where r_i is one of the four roots r_i of the quartic polynomial $x^4 - x^3 - 4x^2 + 4x + 1 = 0$, where $r_1 = -2 \cos(\pi/15) \approx -1.956295$, $r_2 = -2 \cos(11\pi/15) \approx 1.338261$, $r_3 = -2 \cos(7\pi/15) \approx -0.209056$ and $r_4 = -2 \cos(13\pi/15) \approx 1.827091$.

It is known that such a quartic contains the golden ratio $\phi = z_2/2$ and the irrational $\sqrt{2}$ in its structure. Following [24, Section 3], the inflection secant S has segments whose length ratio is ϕ . Also the double tangent, the inflection secant and the straight line passing through the third tangent point and parallel to them separate areas whose ratio is $\sqrt{2}$.

It may be that some other clues in the mystery of the genetic code (and a possible connection to the present approach) are in a concept introduced

MICHEL PLANAT†, RAYMOND ASCHHEIM‡, MARCELO M. AMARAL‡, FANG FANG‡ AND KLEE IRWIN‡

under the name of BoerdijkCoxeter (BC) helix –an helix made of contiguous tetrahedra– more or less at the same time than the discovery of DNA [25]. The concept was revisited to arrive at a BC helix exhibiting a connection to aperiodicity and the golden ratio [26, 27].

From now, we consider the (genus one) hyperelliptic curve \mathcal{C}

$$\mathcal{C} : y^2 = x^4 - x^3 - 4x^2 + 4x + 1$$

as a potential model of many features of DNA double helix.

The characters responsible of the quadratic irrationalities and the roots of the quartic are defined over the cyclotomic fields $\mathbb{Q}(\zeta_5)$ and $\mathbb{Q}(\zeta_{15})$, respectively, where ζ_n is a primitive n -th complex root of unity. Thus we focus on \mathcal{C} with points over such cyclotomic fields.

Points of \mathcal{C} over cyclotomic fields. As usual for elliptic and hyperelliptic curves of genus g , \mathcal{C} is embedded in a weighted projective plane, with weights 1, $g + 1$, and 1, respectively on coordinates x , y and z . Therefore, point triples are such that $(x : y : z) = (\mu x : \mu y : \mu z)$, μ in the field of definition, and the points at infinity take the form $(1 : y : 0)$. Below, the software Magma is used for the calculation of points of \mathcal{C} [23]. For the points of \mathcal{C} , there is a parameter called ‘bound’ that loosely follows the heights of the x -coordinates found by the search algorithm.

Let us start with points on \mathcal{C} with entries over the cyclotomic field $\mathbb{Q}(\zeta_5)$.

For points whose entries are bounded by modulus 1, one gets 8 cases as follows $(1 : \pm 1 : 0)$ (the points at infinity), $(0 : \pm 1 : 1)$, $(1 : \pm 1 : 1)$ and $(2 + \phi : \pm(3 + \phi) : 1)$. For points whose entries are bounded by modulus 2, one gets 10 extra cases $(\pm 2 : -1 : 0)$, $(\pm 2 : 1 : 1)$, $(2\phi : \pm(3 - 4\phi) : 1)$, $(-2(1 + \phi) : \pm(7 + 4\phi) : 1)$ and $(1 - \phi : \pm(4 + \phi) : 1)$. For points whose entries are bounded by modulus 10, one gets 8 extra cases $(4/3 : \pm 1/9 : 1)$, $(6 : \pm 31 : 1)$, $(4/5 : \pm 31/25 : 1)$, $(-1/7 : \pm 29/49 : 1)$.

Let us continue with points on \mathcal{C} with entries over $\mathbb{Q}(\zeta_{15})$. Points with integer entries are as those given over $\mathbb{Q}(\zeta_5)$. In addition, the points whose entries are bounded by modulus 1 are $(r_1 : 0 : 1)$, $(r_2 : 0 : 1)$, $(r_3 : 0 : 1)$, $(r_4 : 0 : 1)$ (where r_i is one of the four roots of the quartic polynomial above), $(2 + \phi : \pm(3 + \phi) : 1)$ [the latter 2 points are on $\mathbb{Q}(\zeta_5)$] The non rational points whose entries are bounded by modulus 2 are $(2\phi : \pm(3 - 4\phi) : 1)$, $(-2(1 + \phi) : \pm(7 + 4\phi) : 1)$ [the latter 2 points are on $\mathbb{Q}(\zeta_5)$] and $(\phi^2 : \pm(1 + \phi^2) : 1)$. For non rational points whose entries are bounded by modulus 7, one gets 8 extra cases and no more up to bound 10. The entries could not be found related to simple irrationalities and are approximated as $(3.02234 : \pm 5.69063 : 1)$, $(0.90536 : \pm 1.12804 : 1)$, $(0.50552 : \pm 1.39140 : 1)$ and $(11.56677 : \pm 125.95399)$.

Over both cyclotomic fields, there are plenty extra points where the entries have higher modulus but we do not list them here.

The group law on the Jacobian J of the hyperelliptic curve \mathcal{C} .

There exists a group law on the Jacobian J of a hyperelliptic curve [28]. Using Magma we provide some results for the operations in $J(\mathcal{C})$ defined over $\mathbb{Q}(\zeta_5)$.

Let focus on the 8 elements a_i of bound 1 in the Jacobian. In the Mumford representation of Jacobian elements, $a_1 = (1, 0, 0)$, $a_2 = (x^2 - 15/8, -x^2 +$

$49/16, 2)$, $a_3 = (x - 4/5, -x^2 - 3/5, 2)$, $a_4 = (x - 4/3, -x^2 + 5/3, 2)$, $a_5 = (x + 1/7, -x^2 - 4/7, 2)$, $a_6 = (x - 1, -x^2 + 2, 2)$, $a_7 = (x - (2 + \phi), -x^2 + 2(1 + \phi), 2)$, $a_8 = (x + 2.25123, -x^2 + 8.03750, 2)$. In Table 8, the inverses or elements a_i are shown. In Table 9, one explicit the sums $a_i + a_j$ between two elements of bound 1 in the Jacobian.

A similar group law exists for \mathcal{C} defined over $\mathbb{Q}(\zeta_{15})$ but it is more difficult to explicit the points and we do not give details here.

6. CONCLUSION

We unveiled part of the machinery of DNA in building the genetic code thanks to the representations of a distinguished finite group of signature $G = \mathbb{Z}_5 \times 2O$, where $2O$ is the binary octahedral group. The group characters κ of G are found to map to the proteigenomic amino acids and this is performed by almost keeping a complete quantum information when the κ 's are seen as magic states, as in quantum computing. Moreover, there exists a unique quartic and a related hyperelliptic curve \mathcal{C} behind the DNA scene which gives sense to the irrationalities found in the character table (and in the real biological world). It is still too early to conclude that DNA calculates on points of \mathcal{C} or on points of the Jacobian $J(\mathcal{C})$ but it is tempting to work in this direction in future papers. It may be that other finite groups and their attached hyperelliptic curve are governing the secondary structure of proteins such that the alpha helix or the DNA packing into chromatin and chromosomes.

REFERENCES

- [1] J. D. Watson and F. H.C. Crick, A structure for deoxyribose nucleic acid, *Nature* **171** 737-738 (1953).
- [2] L. A. Pray, Discovery of DNA Structure and Function: Watson and Crick, *Nature Education* **1** 100 (2008), paper available at the address <https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/>.
- [3] E. Chargaff, Chemical specificity of nucleic acids and mechanism of their enzymatic degradation, *Experientia* **6** 201209 (1950).
- [4] B. Maddox, The double helix and the 'wronged heroine', *Nature* **421** 407-408 (2003).
- [5] G. Gamov, Possible Relation between deoxyribonucleic acid and protein structures, *Nature* **173** 318 (1954).
- [6] F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, General nature of the genetic code for proteins, *Nature*, **192** 12276-32 (1961).
- [7] E. V. Koonin and A. S. Novozhikov, Origin and evolution of the genetic code: the universal enigma, *IUBMB Life* **61** 99-111 (2009).
- [8] C. R. Woese, D. H. Dugre, W. C. Saxinger and S. A. Dugre, The molecular basis of the genetic code, *Proc. Natl Acad Sci* **55** 966-974 (1966).
- [9] S. V. Petoukhov, Matrix genetics, part 2: the degeneracy of the genetic code and the octave algebra with two quasi-real units (the Yin-Yang octave algebra), Preprint0803.3330 [q-bio].
- [10] J. E.M. Hornos, Y. M.M. Hornos, Algebraic model for the evolution of the genetic code, *Phys. Rev. Lett.* **71** 44014404 (1993).
- [11] M. Di Giulio, The origin of the genetic code: theories and their relationships, *Biosystems* **80** 175-184 (2005).
- [12] C. M. Carlevaro, R. M. Irastorza and F. Vericat, Quaternionic representation of the genetic code, *Biosystems* **141** 10-19 (2016).

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡], MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

- [13] F. Antonelli and M. Forger, Symmetry breaking in the genetic code: Finite groups, *Math. Comp. Mod.* **53** 1469-1488 (2011).
- [14] M. Planat and Z. Gedik, Magic informationally complete POVMs with permutations, *R. Soc. open sci.* **4** 170387 (2017).
- [15] M. Planat, R. Aschheim, M. M. Amaral and K. Irwin, Informationally complete characters for quark and lepton mixings, *Symmetry* **12** 1000 (2020).
- [16] <https://en.wikipedia.org/wiki/DNA> (accessed on 1 May 2020).
- [17] D. Harel, R. Unger and J. L. Sussman, Beauty is in the genes of the beholder, *Trends in Bioch. Sc.* **11** 155-156 (1986).
- [18] J. R. Jungck, The Genetic Code as a Periodic Table, *J. Mol. Evol.* **11** 211-224 (1978).
- [19] D. C. Mathew and Z. Luthey-Schulten, On the physical basis of the amino acid polar requirement, *J. Mol. Evol.* **66** 519-528 (2008).
- [20] C. W. Carter Jr, Coding of Class I and II aminoacyl-tRNA synthetases, *Adv Exp Med Biol.* **966** 103148 (2017).
- [21] C. A. Fuchs, On the quantumness of a Hilbert space, *Quant. Inf. Comp.* **4** 467-478 (2004).
- [22] Planat, M. Pauli graphs when the Hilbert space dimension contains a square: Why the Dedekind psi function? *J. Phys. A Math. Theor.* **2011**, 44, 045301.
- [23] Bosma, W.; Cannon, J. J.; Fieker, C.; Steel, A. (eds). *Handbook of Magma functions*, Edition 2.23 (2017), 5914pp (accessed on 1 January 2019).
- [24] H. TR Aude, Notes on quartic curves, *Am. Math. Month.* **56** 165-170 (1949).
- [25] A. H. Boerdijk, Some remarks concerning close-packing of equal spheres, *Philips Res. Rep.* **7** 30 (1952).
- [26] G. Sadler, F. Fang, R. Clawson and K. Irwin, Periodic modification of the Boerdijk-Coxeter helix (tetrahelix), *Mathematics* **7** 1001 (2019).
- [27] Fang Fang, K. Irwin, J. Kovacs and G. Sadler, Cabinet of curiosities: the interesting geometry of the angle $\beta = \arcsin(3\phi - 1)/4$, *Fractal Fract.* **3** 48 (2019).
- [28] C. Costello and K. Lauter, Group law computations on Jacobians of hyperelliptic curves, in T. Johansson and P. Q.Nguyen, editors, EUROCRYPT, volume 7881 of Lecture Notes in Computer Science(Springer, 2013),pp. 194210.

[†] UNIVERSITÉ DE BOURGOGNE/FRANCHE-COMTÉ, INSTITUT FEMTO-ST CNRS UMR 6174, 15 B AVENUE DES MONTBOUCONS, F-25044 BESANÇON, FRANCE.

E-mail address: michel.planat@femto-st.fr

[‡] QUANTUM GRAVITY RESEARCH, LOS ANGELES, CA 90290, USA

E-mail address: raymond@QuantumGravityResearch.org

E-mail address: Klee@quantumgravityresearch.org

E-mail address: Marcelo@quantumgravityresearch.org

E-mail address: Fang@QuantumGravityResearch.org

κ_1	1	1	1	1	1	1	1	1	1	1	1	1
κ_2	1	1	1	1	-1	1	1	1	-1	-1	1	1
κ_3	2	2	-1	2	0	2	2	-1	0	0	2	2
κ_4	2	2	2	2	0	z_1	z_1^*	2	0	0	z_1	z_1^*
κ_5	2	2	2	2	0	z_1^*	z_1	2	0	0	z_1^*	z_1
κ_6	2	-2	-1	0	0	2	2	1	z_4	$-z_4$	-2	-2
κ_7	2	-2	-1	0	0	2	2	1	$-z_4$	z_4	-2	-2
κ_8	2	2	-1	2	0	z_1	z_1^*	-1	0	0	z_1	z_1^*
κ_9	2	2	-1	2	0	z_1^*	z_1	-1	0	0	z_1^*	z_1
κ_{10}	2	2	-1	2	0	z_1	z_1^*	-1	0	0	z_1	z_1^*
κ_{11}	2	2	-1	2	0	z_1^*	z_1	-1	0	0	z_1^*	z_1
κ_{12}	3	3	0	-1	-1	3	3	0	1	1	3	3
κ_{13}	3	3	0	-1	1	3	3	0	-1	-1	3	3
κ_{14}	4	-4	1	0	0	4	4	-1	0	0	-4	-4
κ_{15}	4	-4	-2	0	0	z_2	z_2^*	2	0	0	$-z_2$	$-z_2^*$
κ_{16}	4	-4	-2	0	0	z_2^*	z_2	2	0	0	$-z_2^*$	$-z_2$
κ_{17}	4	-4	1	0	0	z_2	z_2^*	-1	0	0	$-z_2$	$-z_2^*$
κ_{18}	4	-4	1	0	0	z_2	z_2^*	-1	0	0	$-z_2$	$-z_2^*$
κ_{19}	4	-4	1	0	0	z_2^*	z_2	-1	0	0	$-z_2^*$	$-z_2$
κ_{20}	4	-4	1	0	0	z_2^*	z_2	-1	0	0	$-z_2^*$	$-z_2$
κ_{21}	6	6	0	-2	0	z_3	z_3^*	0	0	0	z_3	z_3^*
κ_{22}	6	6	0	-2	0	z_3^*	z_3	0	0	0	z_3^*	z_3
→	1	1	1	1	1	1	1	1	1	1	d^2	Met
→	1	1	1	1	1	1	1	1	1	1	d^2	Trp
→	-1	-1	-1	-1	2	2	-1	-1	-1	-1	d^2	Cys
→	z_1^*	z_1^*	z_1	z_1	z_1	z_1^*	z_1	z_1	z_1^*	z_1^*	d^2	Phe
→	z_1	z_1	z_1^*	z_1^*	z_1^*	z_1	z_1^*	z_1^*	z_1	z_1	d^2	Tyr
→	-1	-1	-1	-1	0	0	1	1	1	1	d^2	His
→	-1	-1	-1	-1	0	0	1	1	1	1	d^2	Gln
→	r_1	r_2	r_3	r_4	z_1	z_1^*	r_3	r_4	r_2	r_1	d^2	Asn
→	r_4	r_3	r_1	r_2	z_1^*	z_1	r_1	r_2	r_3	r_4	d^2	Lys
→	r_2	r_1	r_4	r_3	z_1	z_1^*	r_4	r_3	r_1	r_2	d^2	Glu
→	r_3	r_4	r_2	r_1	z_1^*	z_1	r_2	r_1	r_4	r_3	d^2	Asp
→	0	0	0	0	-1	-1	0	0	0	0	d^2	Ile
→	0	0	0	0	-1	-1	0	0	0	0	475	Stop
→	1	1	1	1	0	0	-1	-1	-1	-1	483	Leu, Pyl, Sec
→	$-z_1$	$-z_1$	$-z_1^*$	$-z_1^*$	0	0	z_1^*	z_1^*	z_1	z_1	480	Leu
→	$-z_1^*$	$-z_1^*$	$-z_1$	$-z_1$	0	0	z_1	z_1	z_1^*	z_1^*	d^2	Val
→	$-r_3$	$-r_4$	$-r_2$	$-r_1$	0	0	r_2	r_1	r_4	r_3	d^2	Pro
→	$-r_4$	$-r_3$	$-r_1$	$-r_2$	0	0	r_1	r_2	r_3	r_4	d^2	Thr
→	$-r_1$	$-r_2$	$-r_3$	$-r_4$	0	0	r_3	r_4	r_2	r_1	d^2	Ala
→	$-r_2$	$-r_1$	$-r_4$	$-r_3$	0	0	r_4	r_3	r_1	r_2	d^2	Gly
→	0	0	0	0	$-z_1$	$-z_1^*$	0	0	0	0	d^2	Ser
→	0	0	0	0	$-z_1^*$	$-z_1$	0	0	0	0	d^2	Arg

TABLE 7. The character table for the group $G = (240, 105) \cong \mathbb{Z}_5 \times 2O$. The two last hand side columns are for the rank of the Gram matrix (with $d = 22$) for the corresponding character and the assignment of an amino acid of the genetic code, respectively (see also Table 3 and 4 for further details). The notation in the entries is as follows: $z_1 = -(\sqrt{5} + 1)/2$, $z_2 = \sqrt{5} - 1$, $z_3 = 3(1 + \sqrt{5})/2$, $z_4 = \sqrt{2}$, the exponent * means the algebraic conjugation, e.g. $z_1^* = (\sqrt{5} - 1)/2$, the r_i 's are the four real roots of the quartic curve $x^4 - x^3 - 4x^2 + 4x + 1 = 0$, i.e. $r_1 \approx -1.956295$, $r_2 \approx 1.338261$, $r_3 \approx -0.209056$ and $r_4 \approx 1.827091$.

MICHEL PLANAT[†], RAYMOND ASCHHEIM[‡], MARCELO M. AMARAL[‡], FANG FANG[‡] AND KLEE IRWIN[‡]

points of J	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
inverse	a_1	4	5	15	7	a_6	a_7	?

TABLE 8. The inverse of the 8 elements a_i of bound 1 in the Jacobian J of the hyperelliptic curve \mathcal{C} over $\mathbb{Q}(\zeta_5)$. One gets $-a_2 = (1, -x^2 + x/2, 2)$, $-a_3 = (x, -x^2 - 1, 2)$, $-a_4 = (x, -x^2 + 1, 2)$ and $-a_5 = (x - 1, -x^2, 2)$ of bounds 4, 5, 15 and 7, respectively.

sum	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
a_1	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
a_2	.	15	5	4	7	a_5	a_8	?
a_3	.	.	b_6	a_2	6	b_8	b_{10}	54
a_4	.	.	.	b_8	b_6	b_1	b_2	14
a_5	15	a_2	b_9	?
a_6	a_1	b_4	b_9
a_7	a_1	a_2
a_8	15

TABLE 9. The addition table between the 8 elements a_i of bound 1 in the Jacobian J of the hyperelliptic curve \mathcal{C} over $\mathbb{Q}(\zeta_5)$. The 10 elements with bound 2 in the Jacobian are denoted b_i . For entries with bound not 1 or 2, only the bound is given. The points of the Jacobian with bounds 2 used in the table are $b_1 = (x + 2, -x^2 + 3, 2)$, $b_6 = (x - 70/31, -x^2 + 221/31, 2)$, $b_8 = (x - 6, -x^2 + 5, 2)$, $b_2 \approx (x + 0.19821, -x^2 + 0.28285, 2)$, $b_9 \approx (x - 1.30757, -x^2 - 2.47363, 2)$ and $b_{10} \approx (x - 2.00889, -x^2 - 1.01292, 2)$.