

**DEPARTAMENTO DE MATEMÁTICA APLICADA**

**Relatório Técnico**

**RT-MAP-0807**

**ON AMINO ACID AND CODON  
ASSIGNMENT IN ALGEBRAIC MODELS  
FOR THE GENETIC CODE**

**Fernando Antoneli, Michael Forger,  
Paola A. Gaviria and  
José Eduardo M. Hornos**

**Agosto, 2008**



**UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA**

**SÃO PAULO — BRASIL**

# On Amino Acid and Codon Assignment in Algebraic Models for the Genetic Code\*

Fernando Antoneli<sup>1</sup>, Michael Forger<sup>1</sup>, Paola A. Gaviria<sup>1</sup>  
and  
José Eduardo M. Hornos<sup>2</sup>

<sup>1</sup> Departamento de Matemática Aplicada  
Instituto de Matemática e Estatística  
Universidade de São Paulo  
BR-05315-970 São Paulo, S.P., Brazil

<sup>2</sup> Departamento de Física e Ciências dos Materiais,  
Instituto de Física de São Carlos,  
Universidade de São Paulo,  
Caixa Postal 369,  
BR-13560-970 São Carlos, S.P., Brazil

## Abstract

We give a list of all possible schemes for performing amino acid and codon assignments in algebraic models for the genetic code which are consistent with a few simple symmetry principles, in accordance with the spirit of the algebraic approach to the evolution of the genetic code proposed by Hornos and Hornos. Our results are complete in the sense of covering all the algebraic models that arise within this approach, whether based on Lie groups/Lie algebras, on Lie superalgebras or on finite groups.

Universidade de São Paulo  
RT-MAP-0807  
August 2008

---

\*Work partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), Brazil

# 1 Introduction

Understanding the origin and primordial evolution of the genetic code is one of the most intriguing problems of modern evolutionary biology, due to its close connection with the question for the origin of life. Over the last few decades, this issue has been approached from many different points of view, as illustrated, for example, by the discussion on the chronological order in which the twenty “biological” aminoacids have been incorporated into the code: there are presently about 40 models referring to various aspects of the early evolution of life on Earth that lead to predictions about this order; see Ref. [1] for an overview. Among them is the algebraic model for the evolution of the genetic code first proposed in 1993 [2] (see also [3,4]) and further developed in subsequent years [5–13].

The present paper is devoted to the analysis of the issue of amino acid and codon assignment in the algebraic approach. The first such assignment was performed in the original paper, based on a best fit to a certain physico-chemical parameter called Grantham polarity [2]. An entirely different approach was formulated in Ref. [5], where it was observed that the algebraic model of Ref. [2] allows for a natural subdivision of codons into 32 “bosonic” ones and 32 “fermionic” ones, according to whether they belong to a multiplet of integral spin or of half-integral spin under a certain  $SU(2)$ -subgroup of the primordial symmetry group  $Sp(6)$ . As it turns out, one can then formulate two simple criteria that allow to fix the codon assignment up to a relatively small number of remaining ambiguities:

- *Family box completeness*: codons in the same family box, that is, codons of the form  $XYN$  where  $N$  runs through the four possible choices  $U, C, A, G$ , are either all “bosonic” or all “fermionic”.
- *Partial Watson-Crick duality*: family boxes dual to each other are either both “bosonic” or both “fermionic”, the dual of a family box  $XYN$  in this sense being defined by  $X^{\dagger}Y^{\dagger}N$  or by  $Y^{\dagger}X^{\dagger}N$ , where  $U^{\dagger} = A, C^{\dagger} = G, A^{\dagger} = U, G^{\dagger} = C$  and  $N$  stands for any of the four possible choices  $U, C, A, G$ .

The corresponding tree of evolution, with an amino acid assignment compatible with these two rules, is shown in Fig. 1. The remaining ambiguities just mentioned result from the option of permuting the amino acids assigned to the five bosonic doublets (Phe, Asp, Glu, Asn, Lys) and four fermionic doublets (Cys, Tyr, Gln, His), to the four fermionic quartets (Gly, Pro, Val, Thr) and to the three sextets (Arg, Leu, Ser) freely among each other, which after excluding trivial permutations leads to a total of  $30 \times 12 \times 6 \times 3 = 12.960$  assignments allowed by these two rules. This compares favorably to the  $3!5!2!9!2! = 1.045.094.400$  possible assignments of the twenty “biological” amino acids to the three sextets, five quartets, two triplets, nine doublets and two singlets of the genetic code that would be allowed if no additional criteria were imposed.

It should be noticed at this point that the analysis in Ref. [5] whose results we have just summarized was carried out exclusively in the context of the specific model of Ref. [2]

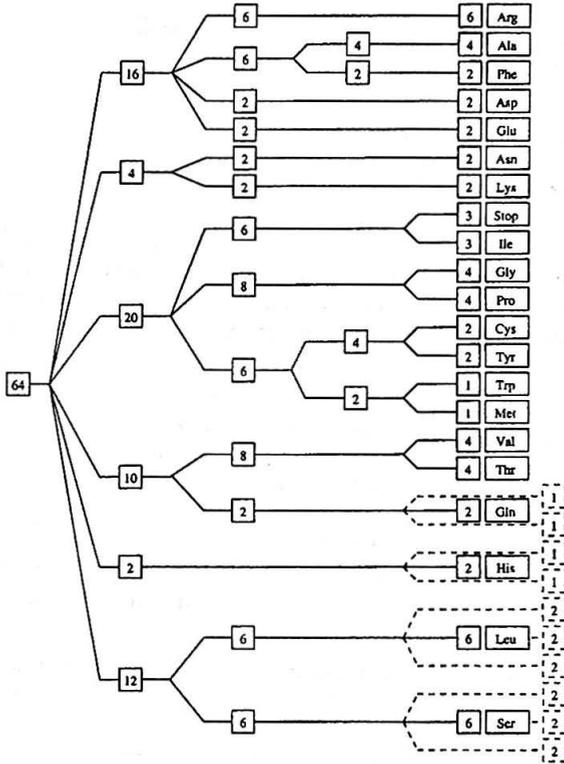


Figure 1: Tree of evolution for the standard genetic code in the  $Sp(6)$  model.

which, in accordance with the primordial symmetry group on which it is based, may be referred to as the  $Sp(6)$ -model. What is missing is a corresponding analysis for the other algebraic models that have meanwhile be found. Indeed, apart from the  $Sp(6)$ -model, there are<sup>1</sup>

- two models based on the simple Lie group  $G_2$  or Lie algebra  $\mathfrak{g}_2$ ,
- six models based on the simple Lie group  $SO(14)$  or Lie algebra  $\mathfrak{so}(14)$  ( $D_7$ ),
- three models based on the simple Lie superalgebra  $\mathfrak{osp}(5|2)$  ( $B(2|1)$ ),

<sup>1</sup>This list has been shown to be complete if we exclude chains starting from the codon representations of the "large" simple Lie groups  $Sp(64)$ ,  $SO(64)$ ,  $SU(64)$  and of the "large" simple finite groups and their satellites, which are the alternating groups  $Alt_{14}$ ,  $Alt_{15}$ ,  $Alt_{65}$  and the symmetric groups  $Sym_{13}$ ,  $Sym_{14}$ ,  $Sym_{65}$ .

- twelve models based on the bicyclic extension  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  of the simple finite group  $B_2(3)$ ,
- one model based on the cyclic extension  $\mathbb{Z}_2.C_3(2)$  of the simple finite group  $C_3(2)$ ,
- six models based on the simple finite group  $G_2(3)$ .

In the case of groups, all of these models are obtained by starting out from (one of) the codon representations ( $\equiv$  64-dimensional irreducible representations) of  $G$ , which is one of the groups listed above and is referred to as the *primordial* symmetry group, and carrying out the procedure of (stepwise) symmetry breaking along an appropriate chain of subgroups

$$G \supset \dots \supset H \supset K, \quad (1)$$

according to the pertinent branching rules. More details will be given in the next section, but for the time being it is sufficient to note that, due to the occurrence of the phenomenon of “freezing” [6], it is important to consider not only the last subgroup  $K$  in the chain, which following common usage we shall refer to as the *residual* symmetry group, but also the penultimate subgroup  $H$ , which we propose to call the *pre-residual* symmetry group. This distinction has originally been proposed in the context of finite groups [10–13], where it is particularly useful, since it allows to interpret frozen multiplets as multiplets under  $H$  and fully broken multiplets as multiplets under  $K$ .

In order to analyze the question of amino acid and codon assignment based on the two criteria spelled out above, the first issue to be addressed is to give an appropriate definition of the “fermion operator” that allows to subdivide the multiplets encountered in a given specific model into “bosons” and “fermions”. To avoid misunderstandings, we would like to stress that this terminology has been borrowed from physics in a purely formal sense, by mere analogy: it is not intended to have any deeper meaning, nor does it bear any relation to the original meaning of the words “boson” and “fermion”, which refers to the statistics of elementary particles. For the purpose of the analysis to be carried out in what follows, a “fermion operator” is simply defined to be any involution ( $\equiv$  element of order 2) in the primordial symmetry group which belongs to the center of one of the intermediate groups in the chain, that is, to the center of any group in the chain except the very first (the primordial symmetry group) and the very last (the residual symmetry group). This implies, in particular, that a fermion operator must belong to the centralizer  $C_G(H)$  of the pre-residual symmetry group  $H$  in  $G$ . Correspondingly, considering the action of the operator that represents this involution in the codon representation, the multiplets encountered in a given specific model can, even in the presence of freezing, be subdivided according to whether they belong to eigenvalue +1 (“bosons”) or  $-1$  (“fermions”). This being understood, we shall in what follows omit the quotation marks.

Analogous terminology will be used when groups are replaced by algebras and the procedure of (stepwise) symmetry breaking is implemented through an appropriate chain

of subalgebras

$$\mathfrak{g} \supset \dots \supset \mathfrak{h} \supset \mathfrak{k}. \quad (2)$$

Note also that our definition of fermion operator, originally formulated for chains of Lie groups or of finite groups, applies equally well to chains of Lie algebras and Lie superalgebras, since it should be understood that for chains of Lie algebras, the right place to look for fermion operators is in the corresponding connected Lie subgroups of the primordial symmetry group, and the same goes for the chains of Lie superalgebras considered here since these employ Lie superalgebras with non-trivial odd part only to describe the primordial symmetry.

## 2 Fermion Operators

In the first step of our analysis, we shall determine, for each of the chains that are known to reproduce the degeneracies of the genetic code, the possible fermion operators.

The simplest case is that of finite groups, since it can be verified by a direct calculation with GAP [14] that for all chains based on finite groups which arise in the genetic code, the centralizer  $C_G(H)$  of  $H$  in  $G$  is already contained in  $H$  and hence equals the center  $Z(H)$  of  $H$ , and that in all cases  $H$  does contain at least one non-trivial central involution. Moreover, the action of such a central involution  $z$  on the space  $V$  of an irreducible representation can be determined from the character table of the group  $H$  since  $z$  acts on  $V$  by  $\pm \text{id}$  where the sign  $\pm$  is given by the value of the corresponding irreducible character on the conjugacy class of  $z$ , which consists of  $z$  alone since  $z$  is central.

For the chains based on Lie algebras or superalgebras, the situation is more complicated since here, the centralizer  $C_G(H)$  of  $H$  in  $G$  is often much larger than the center  $Z(H)$  of  $H$ . Thus, in principle, we must perform a case-by-case analysis to determine all involutions belonging to the center of each of the intermediate groups which appear in the various possible chains. Fortunately, it turns out that it is sufficient to consider only one intermediate group in each chain (or class of chains), namely the one appearing at the end of what in Ref. [7] is called the first phase, which consists in breaking the primordial symmetry down to  $SU(2)$ -symmetries. Correspondingly, this intermediate group is a “maximal  $SU(2)$  type subgroup” in the sense of being a direct product of  $SU(2)$  groups with the maximal possible number of factors.

In order to see why it is sufficient to consider only the central involutions belonging to this “maximal  $SU(2)$  type subgroup”, we note first that no new central involutions arise during posterior steps of the symmetry breaking process, which in Ref. [7] are assembled into a second phase and a third phase. Indeed, during the second phase, which consists in diagonally breaking (some of the)  $SU(2)$ -symmetries, the center shrinks from a direct product of  $\mathbb{Z}_2$ -groups to a direct product of  $\mathbb{Z}_2$ -groups with a smaller number of factors, since the center of the product group  $SU(2) \times SU(2)$  contains the center of its diagonal subgroup  $SU(2)$ . Similarly, during the third phase, which consists in breaking one or

several of the remaining  $SU(2)$  groups either to the full orthogonal group  $O(2)$  or to the special orthogonal group  $SO(2)$ , the center  $\mathbb{Z}_2$  is unaltered in the first case and, although it grows to become equal to all of  $SO(2)$  in the second case, does not acquire any new involutions ( $\equiv$  elements of order 2).

In the opposite direction, the general rule is that the number of central involutions also decreases when we go back from the "maximal  $SU(2)$  type subgroup" to previous intermediate subgroups. As a simple first example, consider the only such inclusion that appears in the  $\mathfrak{sp}(6)$ -chain

$$\mathfrak{sp}(4) \oplus \mathfrak{su}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) ,$$

and similarly in the the  $\mathfrak{osp}(5|2)$ -chains

$$\mathfrak{so}(5) \oplus \mathfrak{sp}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) .$$

Note that this is really the same inclusion, since  $\mathfrak{sp}(2) \cong \mathfrak{so}(3) \cong \mathfrak{su}(2)$  and  $\mathfrak{sp}(4) \cong \mathfrak{so}(5)$ . At the level of simply connected Lie groups, it reads

$$Sp(4) \times SU(2) \supset SU(2) \times SU(2) \times SU(2) ,$$

or equivalently

$$Spin(5) \times Sp(2) \supset SU(2) \times SU(2) \times SU(2) .$$

Obviously, the center  $\mathbb{Z}_2$  of  $Sp(4) \cong Spin(5)$  is contained in the center  $\mathbb{Z}_2 \times \mathbb{Z}_2$  of  $SU(2) \times SU(2)$ .

In order to deal with the same question for the inclusions appearing in the  $\mathfrak{so}(14)$ -chains, which all start out from (one of) the chiral spinor representation(s) of  $\mathfrak{so}(14)$ , we shall adopt a systematic approach based on the theory of Clifford algebras and the spin groups which, unlike the special orthogonal groups, are simply connected and hence allow to accomodate arbitrary representations; see Ref. [15] for the relevant background material. Briefly, the group  $Spin(n)$  is constructed as a subgroup of the group of invertible elements in the even part  $C^+(n)$  of the (real) Clifford algebra  $C(n)$  over  $n$ -dimensional Euclidean space  $\mathbb{R}^n$ . Any (oriented) orthonormal basis of  $\mathbb{R}^n$  provides a set  $\{\gamma_1, \dots, \gamma_n\}$  of generators of  $C(n)$  subject to the basic Dirac anticommutation relations  $\gamma_i \gamma_j + \gamma_j \gamma_i = 2 \delta_{ij}$ , and writing  $\tilde{\gamma} = \gamma_1 \dots \gamma_n$  for their product, we have

- for  $n$  odd, the center of  $C^+(n)$  is  $\mathbb{R}1$  and the center of  $Spin(n)$  is

$$Z(Spin(n)) = \{+1, -1\} \cong \mathbb{Z}_2 ,$$

- for  $n$  even, the center of  $C^+(n)$  is  $\mathbb{R}1 \oplus \mathbb{R}\tilde{\gamma}$  and the center of  $Spin(n)$  is

$$Z(Spin(n)) = \{+1, -1, +\tilde{\gamma}, -\tilde{\gamma}\} \cong \begin{cases} \mathbb{Z}_2 \times \mathbb{Z}_2 & \text{if } n = 0 \pmod{4} (\tilde{\gamma}^2 = +1) \\ \mathbb{Z}_4 & \text{if } n = 2 \pmod{4} (\tilde{\gamma}^2 = -1) \end{cases} .$$

Explicitly, the first few of them are

$$\begin{aligned}
Spin(1) &\cong \mathbb{Z}_2, \\
Spin(2) &\cong U(1), \\
Spin(3) &\cong SU(2), \\
Spin(4) &\cong SU(2) \times SU(2), \\
Spin(5) &\cong Sp(4), \\
Spin(6) &\cong SU(4).
\end{aligned}$$

Now when  $n = p + q$  ( $p, q \geq 1$ ), there is a canonical isomorphism  $C(n) \cong C(p) \hat{\otimes} C(q)$  of Clifford algebras, where  $\hat{\otimes}$  denotes the graded tensor product, which induces the following inclusion between spin groups:

$$Spin(p) \times_{\mathbb{Z}_2} Spin(q) \subset Spin(n).$$

Here the symbol  $\times_{\mathbb{Z}_2}$  indicates taking the direct product  $\times$  and then dividing out by the central  $\mathbb{Z}_2$  subgroup generated by  $(-1, -1)$ . Note that under the respective covering homomorphisms, it corresponds to the following obvious inclusion between special orthogonal groups:

$$SO(p) \times SO(q) \subset SO(n).$$

As far as the centers are concerned, we then have the following situation.

- If  $p$  and  $q$  are odd,  $n$  is even, and the center of  $Spin(n)$  contains two elements,  $\tilde{\gamma}_n$  and  $-\tilde{\gamma}_n$ , which are not contained in the center  $\mathbb{Z}_2 \times_{\mathbb{Z}_2} \mathbb{Z}_2$  of  $Spin(p) \times_{\mathbb{Z}_2} Spin(q)$ .
- If  $p$  is odd and  $q$  is even or  $p$  is even and  $q$  is odd,  $n$  is odd, and the center  $\mathbb{Z}_2$  of  $Spin(n)$  is contained in the center of  $Spin(p) \times_{\mathbb{Z}_2} Spin(q)$ .
- If  $p$  and  $q$  are even,  $n$  is even, and the center of  $Spin(n)$  is contained in the center of  $Spin(p) \times_{\mathbb{Z}_2} Spin(q)$ :  $\tilde{\gamma}_n$  corresponds to  $(\tilde{\gamma}_p, \tilde{\gamma}_q)$  or  $(-\tilde{\gamma}_p, -\tilde{\gamma}_q)$  and  $-\tilde{\gamma}_n$  corresponds to  $(\tilde{\gamma}_p, -\tilde{\gamma}_q)$  or  $(-\tilde{\gamma}_p, \tilde{\gamma}_q)$ .

Analyzing the  $so(14)$ -chains that reproduce the degeneracies of the genetic code, as depicted graphically in Figs 1-4 of Ref. [7], we therefore conclude that there are only a few inclusions where the center of the larger group is larger than that of the smaller group, and all but one are inclusions of the above type with  $p$  and  $q$  both odd. These can be analyzed separately:

- $so(10)$ : this subalgebra appears in chains of the form

$$so(14) \supset su(2) \oplus su(2) \oplus so(10) \supset su(2) \oplus su(2) \oplus \mathfrak{h} \supset \dots$$

where

$$\mathfrak{h} = so(9) \text{ or } su(2) \oplus so(7) \text{ or } sp(4)$$

The corresponding symmetry breaking scheme is shown in the following table.

$D_7$		$(A_1)^2 \oplus D_5$	
HW	$d$	HW	$d$
(0000010)	64	(0,1,00010)	32
		(1,0,00001)	32

Moreover, upon restriction to any of the three subalgebras  $\mathfrak{h}$ , the two chiral spinor representations of  $\mathfrak{so}(10)$  (with highest weights (00010) and (00001)) that appear in this chain remain irreducible and in fact induce the same irreducible representation of  $\mathfrak{h}$  (with highest weight (0001) under  $\mathfrak{so}(9)$ , (1,001) under  $\mathfrak{su}(2) \oplus \mathfrak{so}(7)$  and (11) under  $\mathfrak{sp}(4)$ ). However, they can be distinguished by the value of the central involution of (any one of) the other two  $SU(2)$ -subgroups, so no information is lost by discarding those elements of the center of  $Spin(10)$  that are not contained in the center of the next smaller subgroup.

- $\mathfrak{so}(8)$ : this subalgebra appears in chains of the form

$$\mathfrak{so}(14) \supset \mathfrak{su}(4) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(4) \oplus \mathfrak{h} \supset \dots$$

or

$$\mathfrak{so}(14) \supset \mathfrak{k} \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{h} \supset \dots$$

where

$$\mathfrak{h} = \mathfrak{su}(2) \oplus \mathfrak{sp}(4) \quad \text{or} \quad \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2)$$

and

$$\mathfrak{k} = \mathfrak{su}(4) \oplus \mathfrak{so}(8) \quad \text{or} \quad \mathfrak{su}(2) \oplus \mathfrak{so}(11)$$

or in the chain

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{so}(10) \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{so}(9) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \\ &\supset \dots \end{aligned}$$

The corresponding symmetry breaking schemes are shown in the following tables.

$D_7$		$A_3 \oplus D_4$		$(A_1)^2 \oplus D_4$	
HW	$d$	HW	$d$	HW	$d$
(0000010)	64	(001,0010)	32	(1,1,0010)	32
		(100,0001)	32	(1,1,0001)	32

$D_7$		$A_1 \oplus B_5$		$(A_1)^2 \oplus D_4$	
HW	$d$	HW	$d$	HW	$d$
(0000010)	64	(1,00001)	64	(1,1,0010)	32
				(1,1,0001)	32

$D_7$		$(A_1)^2 \oplus D_5$		$(A_1)^2 \oplus B_4$		$(A_1)^2 \oplus D_4$	
HW	$d$	HW	$d$	HW	$d$	HW	$d$
(0000010)	64	(0,1,00010)	32	(0,1,0001)	32	(0,1,0010)	16
						(0,1,0001)	16
		(1,0,00001)	32	(1,0,0001)	32	(1,0,0010)	16
						(1,0,0001)	16

Moreover, upon restriction to the two subalgebras  $\mathfrak{h}$ , the two chiral spinor representations of  $\mathfrak{so}(8)$  (with highest weights (0010) and (0001)) that appear in all these tables split as shown in the following tables:

$D_4$		$A_1 \oplus C_2$	
HW	$d$	HW	$d$
(0010)	8	(0,01)	5
		(2,00)	3
(0001)	8	(1,10)	8

$D_4$		$(A_1)^4$	
HW	$d$	HW	$d$
(0010)	8	(0,1,1,0)	4
		(1,0,0,1)	4
(0001)	8	(0,1,0,1)	4
		(1,0,1,0)	4

This shows that they can still be distinguished by the value of the central involution of one of the resulting  $SU(2)$ -subgroups, so again no information is lost by discarding those elements of the center of  $Spin(8)$  that are not contained in the center of the next smaller subgroup.

- $\mathfrak{so}(6)$ : this subalgebra appears in chains of the form

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(4) \oplus \mathfrak{so}(8) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{so}(8) \supset \dots \end{aligned}$$

or

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(4) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(4) \oplus \mathfrak{su}(2) \oplus \mathfrak{sp}(4) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{sp}(4) \supset \dots \end{aligned}$$

or

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(4) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(4) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2) \supset \dots \end{aligned}$$

The corresponding symmetry breaking schemes are shown in the following tables.

$D_7$		$A_3 \oplus D_4$		$A_3 \oplus A_1 \oplus C_2$	
HW	$d$	HW	$d$	HW	$d$
(0000010)	64	(001,0010)	32	(001,0,01)	20
				(001,2,00)	12
		(100,0001)	32	(100,1,10)	32

$D_7$		$A_3 \oplus D_4$		$A_3 \oplus (A_1)^4$	
HW	$d$	HW	$d$	HW	$d$
(0000010)	64	(001,0010)	32	(001,0,1,1,0)	16
				(001,1,0,0,1)	16
		(100,0001)	32	(100,0,1,0,1)	16
				(100,1,0,1,0)	16

Moreover, upon restriction to the subalgebra  $\mathfrak{su}(2) \oplus \mathfrak{su}(2)$ , the two chiral spinor representations of  $\mathfrak{so}(6)$  (with highest weights (100) and (001)) that appear in these tables remain irreducible and in fact induce the same irreducible representation of  $\mathfrak{su}(2) \oplus \mathfrak{su}(2)$  (with highest weight (1, 1)). However, they can be distinguished by the value of the central involution of one of the other two  $SU(2)$ -subgroups (that, in case of the first of these chains, appear after also breaking the  $\mathfrak{so}(8)$ ), so again no information is lost by discarding those elements of the center of  $Spin(6)$  that are not contained in the center of the next smaller subgroup.

We are left with a unique case in which the center of the larger group is larger than that of the smaller group, given by the inclusion  $\mathfrak{so}(7) \supset \mathfrak{g}_2$ , since the center of  $Spin(7)$  is  $\mathbb{Z}_2$  while that of  $G_2$  is trivial. However, inspection of the symmetry breaking schemes where  $B_3$  appears reveal that every single one of these schemes contains one and the same irreducible representation of  $\mathfrak{so}(7)$ , namely the spinor representation (with highest weight (001)), on which the non-trivial element of the center  $\mathbb{Z}_2$  of  $Spin(7)$  acts as minus the identity. This implies that in all these schemes, the central involution in  $Spin(7)$  that does not belong to  $G_2$  acts as minus the identity on the entire codon space and hence may be discarded without loss of information.

Having proved that all central involutions that need to be considered may be found in the “maximal  $SU(2)$  type subgroup” mentioned before, we proceed to the enumeration of the possible fermion operators for the chains of subalgebras or subgroups that reproduce the degeneracies of the genetic code. In each of these chains, we indicate this subgroup by framing the corresponding subalgebra, which is the direct sum of a certain number, say  $k$ , of copies of  $\mathfrak{su}(2)$ , leading to a total number of  $2^k - 1$  non-trivial central involutions. Depending on the specific codon representation under consideration, further reductions in the number of central involutions that need to be analyzed can be achieved by observing which of them act in the same way or are even irrelevant since they act as a multiple of the identity. As a result, they lead to relatively few possible schemes for classifying part or all of the codons as bosonic or fermionic. Most of these schemes are balanced in the sense that they contain an equal number of bosonic and fermionic family boxes (8 each); in particular, all schemes derived from Lie algebras or Lie superalgebras are of this kind, exhibiting 32 bosonic and 32 fermionic codons. In the case of schemes derived from finite groups, however, we also encounter unbalanced schemes, in which the number of bosonic and of fermionic codons is different. All these schemes are presented in Tables 1 and 2 at the end of this section.

(1) The  $\mathfrak{sp}(6)$ -chain (eq. (7) of Ref. [7], p. 3189):

$$\mathfrak{sp}(6) \supset \mathfrak{sp}(4) \oplus \mathfrak{su}(2) \supset \boxed{\mathfrak{su}(2)^3} \supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{su}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{so}(2)$$

There are 7 non-trivial central involutions in  $SU(2)^3$  among which only three need to be considered, since  $(-1, -1, -1)$  acts as minus the identity on the entire codon space; they can be chosen as  $(-1, 1, 1)$ ,  $(1, -1, 1)$  and  $(1, 1, -1)$ , say. Inspection of the corresponding symmetry breaking scheme (Table 4 of Ref. [7], p. 3189), combined with the criterion of family box completeness, leads to the following conclusions:

- (i) The action of  $(-1, 1, 1)$  provides the scheme shown in item (a) of Table 1 below, where the lightly shaded family boxes must contain 4 bosonic and 2 fermionic doublets.
- (ii) The action of  $(1, -1, 1)$  provides the scheme shown in item (b) of Table 1 below, where the lightly shaded family boxes must contain 1 bosonic and 4 fermionic quartets as well as 4 bosonic and 2 fermionic doublets.
- (iii) The action of  $(1, 1, -1)$  provides the scheme shown in item (c) of Table 1 below, where the lightly shaded family boxes must contain 1 bosonic and 4 fermionic quartets as well as 4 bosonic and 2 fermionic doublets.

(2) The  $\mathfrak{g}_2$ -chains (eqs (8) & (9) of Ref. [7], pp. 3190 & 3191):

$$\begin{aligned} \mathfrak{g}_2 &\supset \boxed{\mathfrak{su}(2)^2} \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{so}(2) \oplus \mathfrak{o}(2) \quad (\text{first chain, eq. (8)}) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{so}(2) \quad (\text{second chain, eq. (9)}) \end{aligned}$$

There are 3 non-trivial central involutions in  $SU(2)^2$  among which only one need to be considered, since  $(-1, -1)$  acts as the identity on the entire codon space; it can be chosen as  $(-1, 1)$ , say. Inspection of the corresponding symmetry breaking scheme (Tables 5 & 6 of Ref. [7], pp. 3190 & 3191) shows that its action is compatible with and in fact completely fixed by the requirement of family box completeness; the result is the same for both chains and is shown in item (d) of Table 1 below.

(3) The first  $\mathfrak{so}(14)$ -chain (eq. (10) of Ref. [7], pp. 3191-3193):

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(4) \oplus \mathfrak{so}(8) \supset \mathfrak{su}(4) \oplus \mathfrak{su}(2) \oplus \mathfrak{sp}(4) \supset \boxed{\mathfrak{su}(2)^5} \supset \mathfrak{su}(2)^4 \supset \mathfrak{su}(2)^3 \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{su}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{so}(2) \oplus \mathfrak{su}(2) \end{aligned}$$

There are 31 non-trivial central involutions in  $SU(2)^5$  among which only three need to be considered, since  $(-1, -1, -1, -1, -1)$  acts as the identity while  $(-1, 1, 1, 1, 1)$  and  $(1, -1, 1, 1, 1)$  act as minus the identity on the entire codon space; they can be chosen as  $(1, 1, -1, 1, 1)$ ,  $(1, 1, 1, -1, 1)$  and  $(1, 1, 1, 1, -1)$ , say. Inspection of the corresponding symmetry breaking scheme (Tables 7 & 8 of Ref. [7], p. 3192) shows that the first two of these are inconsistent with the requirement of family box completeness

since they force the triplets and singlets to be bosonic but eight of the nine doublets to be fermionic; only the third leads to an action that is compatible with and in fact completely fixed by the requirement of family box completeness; the result is shown in item (e) of Table 1 below.

(4) The second  $so(14)$ -chain (eq. (11) of Ref. [7], pp. 3193-3196):

$$\begin{aligned}
so(14) &\supset su(4) \oplus so(8) \supset su(4) \oplus su(2) \oplus sp(4) \supset \boxed{su(2)^5} \supset su(2)^4 \supset su(2)^3 \\
&\supset o(2) \oplus su(2) \oplus su(2) \supset o(2) \oplus su(2) \oplus so(2) \supset so(2) \oplus su(2) \oplus so(2) \\
&\quad \text{(first option)} \\
&\supset o(2) \oplus su(2) \oplus su(2) \supset o(2) \oplus su(2) \oplus so(2) \supset o(2) \oplus so(2) \oplus so(2) \\
&\quad \text{(second option)}
\end{aligned}$$

There are 31 non-trivial central involutions in  $SU(2)^5$  among which only three need to be considered, since  $(-1, -1, -1, -1, -1)$ , acts as the identity while  $(-1, 1, 1, 1, 1)$  and  $(1, -1, 1, 1, 1)$  act as minus the identity on the entire codon space; they can be chosen as  $(1, 1, -1, 1, 1)$ ,  $(1, 1, 1, -1, 1)$  and  $(1, 1, 1, 1, -1)$ , say. Inspection of the corresponding symmetry breaking scheme (Tables 9-11 of Ref. [7], pp. 3194-3196) shows that combining their action with the criterion of family box completeness leads to the following conclusions, for both options:

- (i) The action of  $(1, 1, -1, 1, 1)$ , which corresponds to the action of the central involution  $(1, 1, -1)$  in  $SU(2)^3$ , provides the scheme shown in item (f) of Table 1 below, where the lightly shaded family boxes must contain 3 bosonic and 2 fermionic quartets.
- (ii) The action of  $(1, 1, -1, 1, -1)$ , which corresponds to the action of the central involution  $(1, -1, 1)$  in  $SU(2)^3$ , provides the scheme shown in item (d) of Table 1 below.
- (iii) The action of  $(1, 1, -1, -1, 1)$ , which corresponds to the action of the central involution  $(-1, 1, 1)$  in  $SU(2)^3$ , provides the scheme shown in item (b) of Table 1 below, where the lightly shaded family boxes must contain 2 bosonic and 3 fermionic quartets as well as 2 bosonic and 4 fermionic doublets.

(5) The third  $so(14)$ -chain (eq. (12) of Ref. [7], pp. 3197-3198):

$$\begin{aligned}
so(14) &\supset su(4) \oplus so(8) \supset su(4) \oplus su(2)^4 \supset \boxed{su(2)^6} \supset su(2)^5 \supset su(2)^4 \\
&\supset su(2) \oplus o(2) \oplus su(2) \oplus su(2) \supset su(2) \oplus o(2) \oplus su(2) \oplus so(2) \\
&\supset su(2) \oplus so(2) \oplus su(2) \oplus so(2)
\end{aligned}$$

There are 63 non-trivial central involutions in  $SU(2)^6$  among which only three need to be considered, since  $(-1, 1, 1, 1, 1, 1)$ ,  $(1, -1, 1, 1, 1, 1)$ ,  $(1, 1, -1, -1, 1, 1)$  and  $(1, 1, 1, 1, -1, -1)$  all act as minus the identity on the entire codon space; they can be chosen as  $(1, 1, 1, -1, 1, 1)$ ,  $(1, 1, 1, 1, 1, -1)$  and  $(1, 1, 1, -1, 1, -1)$ , say. Inspection

of the corresponding symmetry breaking scheme (Tables 12 & 13 of Ref. [7], p. 3198) shows that combining their action with the criterion of family box completeness leads to the following conclusions:

- (i) The action of  $(1, 1, 1, -1, 1, 1)$ , which corresponds to the action of the central involutions  $(1, -1, 1, 1)$  and  $(1, 1, -1, 1)$  in  $SU(2)^4$ , provide the scheme shown in item (e) of Table 1 below.
  - (ii) The action of  $(1, 1, 1, 1, 1, -1)$ , which corresponds to the action of the central involutions  $(-1, 1, 1, 1)$  and  $(1, 1, 1, -1)$  in  $SU(2)^4$ , are inconsistent with the requirement of family box completeness since they force the triplets and singlets to be bosonic but eight of the nine doublets to be fermionic.
  - (iii) The action of  $(1, 1, 1, -1, 1, -1)$ , which corresponds to the action of the central involutions  $(-1, -1, 1, 1)$  and  $(1, 1, -1, -1)$  in  $SU(2)^4$ , are inconsistent with the requirement of family box completeness since they force the triplets and singlets to be bosonic but eight of the nine doublets to be fermionic.
- (6) The fourth  $\mathfrak{so}(14)$ -chain (eq. (13) of Ref. [7], pp. 3199-3202):

$$\begin{aligned} \mathfrak{so}(14) &\supset \mathfrak{su}(2) \oplus \mathfrak{so}(11) \supset \mathfrak{su}(2)^3 \oplus \mathfrak{so}(7) \supset \mathfrak{su}(2)^3 \oplus \mathfrak{g}_2 \supset \boxed{\mathfrak{su}(2)^4} \supset \mathfrak{su}(2)^3 \supset \mathfrak{su}(2)^2 \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{so}(2) \oplus \mathfrak{o}(2) \quad (\text{first option}) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{so}(2) \quad (\text{second option}) \end{aligned}$$

There are 15 non-trivial central involutions in  $SU(2)^4$  among which only one need to be considered, since  $(-1, 1, 1, 1)$  and  $(1, -1, -1, 1)$  act as minus the identity while  $(1, 1, 1, -1)$  acts as the identity on the entire codon space; it can be chosen as  $(1, -1, 1, 1)$ , say, whose action corresponds to that of the central involutions  $(-1, 1)$  and  $(1, -1)$  in  $SU(2)^2$ . Inspection of the corresponding symmetry breaking scheme (Tables 14-16 of Ref. [7], pp. 3199-3201) shows that its action is compatible with and in fact completely fixed by the requirement of family box completeness; the result is the same for both options and is shown in item (d) of Table 1 below.

- (7) The  $\mathfrak{osp}(5|2)$ -chain (Tables V-VIII of Ref. [9], pp. 5436 & 5438-5440):

$$\begin{aligned} \mathfrak{osp}(5|2) &\supset \mathfrak{so}(5) \oplus \mathfrak{sp}(2) \supset \boxed{\mathfrak{su}(2)^3} \supset \mathfrak{su}(2)^2 \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{o}(2) \oplus \mathfrak{o}(2) \quad (\text{first option}) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{o}(2) \oplus \mathfrak{o}(2) \quad (\text{second option}) \\ &\supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{o}(2) \oplus \mathfrak{o}(2) \quad (\text{third option}) \end{aligned}$$

There are 7 non-trivial central involutions in  $SU(2)^3$  among which only three need to be considered, since  $(1, -1, -1)$  acts as minus the identity on the entire codon space; they can be chosen as  $(1, 1, -1)$ ,  $(-1, -1, 1)$  and  $(-1, -1, -1)$ . Inspection of the corresponding symmetry breaking schemes (Tables V-VIII of Ref. [9], pp. 5436 & 5438-5440), combined with the criterion of family box completeness, leads to the following conclusions:

- (i) The action of  $(1, 1, -1)$ , which corresponds to the action of the central involution  $(1, -1)$  in  $SU(2)^2$ ,
  - provides the scheme shown in item (b) of Table 1 below, where the lightly shaded family boxes must contain 1 bosonic and 4 fermionic quartets as well as 4 bosonic and 2 fermionic doublets, for the first option,
  - provides the scheme shown in item (f) of Table 1 below, where the lightly shaded family boxes must contain 3 bosonic and 2 fermionic quartets, for the second and third option.
- (ii) The action of  $(-1, -1, 1)$ , which corresponds to the action of the central involution  $(-1, 1)$  in  $SU(2)^2$ , provides the scheme shown in item (d) of Table 1 below, for all three options.
- (iii) The action of  $(-1, -1, -1)$ 
  - is inconsistent with the requirement of family box completeness since it forces the triplets and singlets to be bosonic but eight of the nine doublets to be fermionic, for the first option,
  - provides the scheme shown in item (b) of Table 1 below, where the lightly shaded family boxes must contain 2 bosonic and 3 fermionic quartets as well as 2 bosonic and 4 fermionic doublets, for the second and third option.

(8) The  $B_2(3)$ -chain [10, 13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2^2) \supset Q_8:(3^2 \times 2)$$

Recall that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations which, in this chain, produce two non-equivalent patterns of symmetry breaking: one is obtained from any of the two real representations and the other from any of the two complex representations, which form a complex conjugate pair. Moreover, the intermediate subgroup  $Q_8:(3^2:2^2)$  has three non-trivial central involutions, denoted by  $2a$ ,  $2b$  and  $2c$ . Inspection of the relevant character table shows that (except for a total sign in the last two cases) their action is the same for both patterns and that:

- (i) The involution  $2a$  provides the scheme shown in item (d) of Table 1 below.
  - (ii) The involution  $2b$  provides the slightly unbalanced scheme shown in item (a) of Table 2 below, where the lightly shaded family boxes must contain 2 bosonic doublets and 4 fermionic doublets.
  - (iii) The involution  $2c$  provides the highly unbalanced scheme shown in item (c) of Table 2 below, where only 2 doublets are fermionic and all other multiplets are bosonic.
- (9) The  $B_2(3)$ -chain [10, 13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2^2) \supset Q_8:D_{12}$$

Recall that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations which, in this chain, produce two non-equivalent patterns of symmetry breaking: one is obtained from any of the two real representations and the other from any of the two complex representations, which form a complex conjugate pair. Moreover, the intermediate subgroup  $Q_8:(3^2:2^2)$  has three non-trivial central involutions, denoted by  $2a$ ,  $2b$  and  $2c$ . Inspection of the relevant character table shows that:

(i) The involution  $2a$  provides the scheme shown in item (d) of Table 1 below, for the real representations as well as the complex representations.

(ii) The involution  $2b$

- provides the scheme shown in item (d) of Table 1 below, for the real representations,
- provides (except for a total sign) the slightly unbalanced scheme shown in item (b) of Table 2 below, where the lightly shaded family boxes must contain 1 bosonic and 4 fermionic quartets, for the complex representations.

(iii) The involution  $2c$

- acts trivially (as the identity) on the entire codon space, for the real representations,
- provides (except for a total sign) the highly unbalanced scheme shown in item (d) of Table 2 below, where only 1 quartet is fermionic and all other multiplets are bosonic, for the complex representations.

(10) The  $B_2(3)$ -chain [10,13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2^2) \supset 3^2:2^3$$

Recall that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations which, in this chain, produce two non-equivalent patterns of symmetry breaking: one is obtained from any of the two real representations and the other from any of the two complex representations, which form a complex conjugate pair. Moreover, the intermediate subgroup  $Q_8:(3^2:2^2)$  has three non-trivial central involutions, denoted by  $2a$ ,  $2b$  and  $2c$ . Inspection of the relevant character table shows that:

(i) The involutions  $2a$  and  $2b$  produce schemes which are inconsistent with the requirement of family box completeness since they force the triplets and singlets to be of opposite nature, for the real representations as well as the complex representations.

(ii) The involution  $2c$

- provides the highly unbalanced scheme shown in item (c) of Table 2 below, where only 2 doublets are fermionic and all other multiplets are bosonic, for the real representations,
- acts trivially (as minus the identity) on the entire codon space, for the complex representations.

(11) The  $B_2(3)$ -chain [10, 13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2) \supset Q_8:3^2$$

Recall that there are really two chains of this type and also that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations, but all of these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action provides the scheme shown in item (d) of Table 1 below.

(12) The  $B_2(3)$ -chain [10, 13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2) \supset Q_8:D_6$$

Recall that there are really two chains of this type and also that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations, but all of these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action provides the scheme shown in item (d) of Table 1 below.

(13) The  $B_2(3)$ -chain [10, 13]

$$\mathbb{Z}_2.B_2(3):\mathbb{Z}_2 \supset Q_8:(3^2:2) \supset 3^2:2^2$$

Recall that there are really two chains of this type and also that  $\mathbb{Z}_2.B_2(3):\mathbb{Z}_2$  has four different codon representations, but all of these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action is inconsistent with the requirement of family box completeness since it forces the triplets and singlets to be of opposite nature.

(14) The  $C_3(2)$ -chain [10, 13]

$$\mathbb{Z}_2.C_3(2) \supset (Q_8:3)^2 \supset Q_8:(3^2 \times 2)$$

Recall that  $\mathbb{Z}_2.C_3(2)$  has two (complex conjugate) codon representations, both of which produce the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $(Q_8:3)^2$  has three non-trivial central involutions, denoted by  $2a$ ,  $2b$  and  $2c$ . Inspection of the relevant character table shows that:

- (i) The involution  $2a$  acts trivially (i.e., as minus the identity) on the entire codon space.
- (ii) The involution  $2b$  provides the scheme shown in item (d) of Table 1 below.
- (iii) The involution  $2c$  provides (except for a total sign) the scheme shown in item (d) of Table 1 below.

(15) The  $G_2(3)$ -chain [10, 13]

$$G_2(3) \supset Q_8:(3^2:2) \supset Q_8:3^2$$

Recall that there are really two chains of this type and also that  $G_2(3)$  has two (complex conjugate) codon representations, but these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action provides the scheme shown in item (d) of Table 1 below.

(16) The  $G_2(3)$ -chain [10, 13]

$$G_2(3) \supset Q_8:(3^2:2) \supset Q_8:D_6$$

Recall that there are really two chains of this type and also that  $G_2(3)$  has two (complex conjugate) codon representations, but these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action provides the scheme shown in item (d) of Table 1 below.

(17) The  $G_2(3)$ -chain [10, 13]

$$G_2(3) \supset Q_8:(3^2:2) \supset 3^2:3^2$$

Recall that there are really two chains of this type and also that  $G_2(3)$  has two (complex conjugate) codon representations, but these produce, in both of these chains, the same pattern of symmetry breaking. Moreover, the intermediate subgroup  $Q_8:(3^2:2)$  has only one non-trivial central involution, denoted by  $2a$ . Its action is inconsistent with the requirement of family box completeness since it forces the triplets and singlets to be of opposite nature.

The results are summarized in Table 1 and Table 2 below, where the distribution of codons within the lightly shaded family boxes is as follows:

Table 1, (a): 4 bosonic doublets in 2 shaded family boxes,  
2 fermionic doublets in 1 unshaded family box.

Table 1, (b<sub>1</sub>): 1 bosonic quartet in 1 shaded family box,  
4 fermionic quartets in 4 unshaded family boxes,  
4 bosonic doublets in 2 shaded family boxes,  
2 fermionic doublets in 1 unshaded family box.

Table 1, (b<sub>2</sub>): 2 bosonic quartets in 2 shaded family boxes,  
3 fermionic quartets in 3 unshaded family boxes,  
2 bosonic doublets in 1 shaded family box,  
4 fermionic doublets in 2 unshaded family boxes.

Table 1, (c): 1 bosonic quartet in 1 shaded family box,  
 4 fermionic quartets in 4 unshaded family boxes,  
 4 bosonic doublets in 2 shaded family boxes,  
 2 fermionic doublets in 1 unshaded family box.

Table 1, (f): 3 bosonic quartets in 3 shaded family boxes,  
 2 fermionic quartets in 2 unshaded family boxes.

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(a)

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(b)

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(c)

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(d)

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(e)

first base	second base				third base
U	C	A	G	U	
U	Ph Ph Le Le	Se Se Se Se	Tr Tr TERM TERM	Cy Cy TERM TERM	U C A G
C	Le Le Le Le	Ph Ph Ph Ph	He He Gh Gh	Arg Arg Arg Arg	U C A G
A	Tr Tr Tr Tr	Th Th Th Th	Ass Ass Ass Ass	Se Se Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Arg Arg Glu Glu	Gly Gly Gly Gly	U C A G

(f)

Table 1: Action of fermion operators leading to balanced schemes, with 32 bosonic and 32 fermionic codons. Bosonic families are shaded and fermionic families are unshaded. Light shading indicates undefined family boxes. Regarding the distribution of codons in these boxes between bosonic and fermionic, see text.



### 3 Watson-Crick Duality

Proceeding to the second step of our analysis, we want to determine, for each of the schemes shown in Table 1 and Table 2, whether there is some kind of natural duality transformation on family boxes that leaves them invariant. We shall require such a duality transformation to be induced by a duality transformation

$$X \rightarrow X^\dagger$$

on the individual nucleic bases, which can be thought of as an involutive permutation of the set  $\{U, C, A, G\}$ . A brief look at the permutation group  $\text{Sym}_4$  shows that it contains three full permutations of order 2, with no fixed points, namely

1. Watson-Crick duality:  $U^\dagger = A, C^\dagger = G, A^\dagger = U, G^\dagger = C$
2. Pyrimidine-purine switch:  $U^\dagger = C, C^\dagger = U, A^\dagger = G, G^\dagger = A$
3. Wrong-way Watson-Crick duality:  $U^\dagger = G, C^\dagger = A, A^\dagger = C, G^\dagger = U$

and another six partial permutations of order 2, with two fixed points, namely

4.  $C - G$  pair switch:  $U^\dagger = U, C^\dagger = G, A^\dagger = A, G^\dagger = C$
5.  $U - A$  pair switch:  $U^\dagger = A, C^\dagger = C, A^\dagger = U, G^\dagger = G$
6. Pyrimidine switch ( $U - C$  switch):  $U^\dagger = C, C^\dagger = U, A^\dagger = A, G^\dagger = G$
7. Purine switch ( $A - G$  switch):  $U^\dagger = U, C^\dagger = C, A^\dagger = G, G^\dagger = A$
8.  $C - A$  switch:  $U^\dagger = U, C^\dagger = A, A^\dagger = C, G^\dagger = G$
9.  $U - G$  switch:  $U^\dagger = G, C^\dagger = C, A^\dagger = A, G^\dagger = U$

We do not consider the latter to be natural: they are only mentioned for completeness but will not be analyzed any further.

For the induced action of any of these duality transformations on family boxes of codons, we consider the following two options:

- (i) without inversion of order:  $(XYN)^\dagger = X^\dagger Y^\dagger N$
- (ii) with inversion of order:  $(XYN)^\dagger = Y^\dagger X^\dagger N$

Regarding the first option, we note that it is inconsistent with the schemes shown in Table 2 since these all contain an odd number of bosonic as well as fermionic family boxes (9 bosonic and 7 fermionic for the slightly unbalanced schemes shown in items (a) and (b), 15 bosonic and 1 fermionic for the highly unbalanced schemes shown in items (c) and (d)), so consistency would require the corresponding duality operator to have at least one family box as a fixed point, which is not the case for any of the operators in the first option.

1. Watson-Crick duality ( $U^\dagger = A$ ,  $C^\dagger = G$ ,  $A^\dagger = U$ ,  $G^\dagger = C$ )

(i) option (i) (without inversion of order)

- Table 1, (a): pairs  $UGN$  (bosonic) with  $ACN$  (fermionic): contradiction.
- Table 1, (b): since the quartet  $ACN$  pairs with  $UGN$  and the doublets in  $AAN$  and  $GAN$  pair with  $UUN$  and  $CUN$ , respectively, and hence must all be bosonic, whereas the quartet  $GCN$  pairs with  $CGN$  and hence must be fermionic, it is seen that a consistent solution (with 1 bosonic and 4 fermionic quartets and 4 bosonic and 2 fermionic doublets in the lightly shaded family boxes) emerges if all quartets except  $ACN$  and the doublets in  $CAN$  are chosen to be fermionic; the corresponding coloring is shown in item (a) of Table 3 below.
- Table 1, (c): since the quartet  $GCN$  pairs with  $CGN$  and the doublets in  $AAN$  and  $GAN$  pair with  $UUN$  and  $CUN$ , respectively, and hence must all be bosonic, whereas the quartet  $ACN$  pairs with  $UGN$  and hence must be fermionic, it is seen that a consistent solution (with 1 bosonic and 4 fermionic quartets and 4 bosonic and 2 fermionic doublets in the lightly shaded family boxes) emerges if all quartets except  $GCN$  and the doublets in  $CAN$  are chosen to be fermionic; the corresponding coloring is shown in item (b) of Table 3 below.
- Table 1, (d): pairs  $UUN$  (bosonic) with  $AAN$  (fermionic): contradiction.
- Table 1, (e): pairs  $UGN$  (bosonic) with  $ACN$  (fermionic): contradiction.
- Table 1, (f): pairs  $UUN$  (bosonic) with  $AAN$  (fermionic): contradiction.

(ii) option (ii) (with inversion of order)

- Table 1, (a): pairs  $CUN$  (fermionic) with  $AGN$  (bosonic): contradiction.
- Table 1, (b): pairs  $CUN$  (fermionic) with  $AGN$  (bosonic): contradiction.
- Table 1, (c): since the quartets  $CCN$  and  $GGN$  as well as the quartets  $ACN$  and  $GUN$  pair among themselves while the quartet  $GCN$  remains invariant, whereas the doublets in  $AAN$  and  $GAN$  pair with  $UUN$  and  $UCN$ , respectively, and hence must be bosonic while the doublets in  $CAN$  pair with  $UGN$  and hence must be fermionic, it is seen that a consistent solution (with 1 bosonic and 4 fermionic quartets and 4 bosonic and 2 fermionic doublets in the lightly shaded family boxes) emerges if all quartets except  $GCN$  and the doublets in  $CAN$  are chosen to be fermionic; again, the corresponding coloring is shown in item (b) of Table 3 below.
- Table 1, (d): pairs  $UUN$  (bosonic) with  $AAN$  (fermionic): contradiction.
- Table 1, (e): pairs  $UCN$  (fermionic) with  $GAN$  (bosonic): contradiction.
- Table 1, (f): pairs  $UUN$  (bosonic) with  $AAN$  (fermionic): contradiction.
- Table 2, (a): since the doublets in  $CAN$ ,  $AAN$  and  $GAN$  pair with  $UGN$ ,  $UUN$  and  $UCN$ , respectively, and hence must all be bosonic, we arrive

at a contradiction with the condition that the lightly shaded family boxes must contain 2 bosonic and 4 fermionic doublets.

- Table 2, (b): pairs  $UUN$  (bosonic) with  $AAN$  (fermionic): contradiction.
- Table 2, (c): since none of the doublets in  $CAN$ ,  $AAN$  and  $GAN$  is fixed under the duality operation, we arrive at a contradiction.
- Table 2, (d): taking the quartet  $GCN$ , which is fixed under the duality operation, to be the unique fermionic family box, we arrive at a solution.

## 2. Pyrimidine-purine switch ( $U^\dagger = C$ , $C^\dagger = U$ , $A^\dagger = G$ , $G^\dagger = A$ )

### (i) option (i) (without inversion of order)

- Table 1, (a): pairs  $UCN$  (bosonic) with  $CUN$  (fermionic): contradiction.
- Table 1, (b): pairs  $UCN$  (fermionic) with  $CUN$  (bosonic): contradiction.
- Table 1, (c): pairs  $UAN$  (fermionic) with  $CGN$  (bosonic): contradiction.
- Table 1, (d): pairs  $UUN$  (bosonic) with  $CCN$  (fermionic): contradiction.
- Table 1, (e): pairs  $UUN$  (bosonic) with  $CCN$  (fermionic): contradiction.
- Table 1, (f): pairs  $UCN$  (fermionic) with  $CUN$  (bosonic): contradiction.

### (ii) option (ii) (with inversion of order)

- Table 1, (a): pairs  $UAN$  (bosonic) with  $GCN$  (fermionic): contradiction.
- Table 1, (b): pairs  $CGN$  (fermionic) with  $AUN$  (bosonic): contradiction.
- Table 1, (c): pairs  $CGN$  (bosonic) with  $AUN$  (fermionic): contradiction.
- Table 1, (d): pairs  $UUN$  (bosonic) with  $CCN$  (fermionic): contradiction.
- Table 1, (e): pairs  $UUN$  (bosonic) with  $CCN$  (fermionic): contradiction.
- Table 1, (f): pairs  $CGN$  (fermionic) with  $AUN$  (bosonic): contradiction.
- Table 2, (a): pairs  $UUN$  (bosonic) with  $CCN$  (fermionic): contradiction.
- Table 2, (b): since the quartets  $CCN$ ,  $ACN$ ,  $GUN$ ,  $GCN$  and  $GGN$  pair with  $UUN$ ,  $UGN$ ,  $CAN$ ,  $UAN$  and  $AAN$ , respectively, we arrive at a contradiction with the condition that the lightly shaded family boxes must contain 1 bosonic and 4 fermionic quartets.
- Table 2, (c): taking the doublets in  $GAN$ , which is fixed under the duality operation, to be the unique fermionic family box, we arrive at a solution.
- Table 2, (d): since none of the quartets  $CCN$ ,  $ACN$ ,  $GUN$ ,  $GCN$  and  $GGN$  is fixed under the duality operation, we arrive at a contradiction.

## 3. Wrong-way Watson-Crick duality ( $U^\dagger = G$ , $C^\dagger = A$ , $A^\dagger = C$ , $G^\dagger = U$ )

### (i) option (i) (without inversion of order)

- Table 1, (a): pairs  $UAN$  (bosonic) with  $GCN$  (fermionic): contradiction.
- Table 1, (b): pairs  $CUN$  (bosonic) with  $AGN$  (fermionic): contradiction.
- Table 1, (c): pairs  $CGN$  (bosonic) with  $AUN$  (fermionic): contradiction.

- Table 1, (d): pairs  $UUN$  (bosonic) with  $GGN$  (fermionic): contradiction.
  - Table 1, (e): pairs  $UUN$  (bosonic) with  $GGN$  (fermionic): contradiction.
  - Table 1, (f): pairs  $CUN$  (bosonic) with  $AGN$  (fermionic): contradiction.
- (ii) option (ii) (with inversion of order)
- Table 1, (a): pairs  $AUN$  (bosonic) with  $GCN$  (fermionic): contradiction.
  - Table 1, (b): pairs  $UAN$  (bosonic) with  $CGN$  (fermionic): contradiction.
  - Table 1, (c): pairs  $UAN$  (fermionic) with  $CGN$  (bosonic): contradiction.
  - Table 1, (d): pairs  $UUN$  (bosonic) with  $GGN$  (fermionic): contradiction.
  - Table 1, (e): pairs  $UUN$  (bosonic) with  $GGN$  (fermionic): contradiction.
  - Table 1, (f): pairs  $UAN$  (bosonic) with  $CGN$  (fermionic): contradiction.
  - Table 2, (a): pairs  $UUN$  (bosonic) with  $GGN$  (fermionic): contradiction.
  - Table 2, (b): pairs  $CUN$  (bosonic) with  $GAN$  (fermionic): contradiction.
  - Table 2, (c): taking the doublets in  $CAN$ , which is fixed under the duality operation, to be the unique fermionic family box, we arrive at a solution.
  - Table 2, (d): taking either the quartet  $ACN$  or the quartet  $GUN$ , which are both fixed under the duality operation, to be the unique fermionic family box, we arrive at a solution.

first base	second base				third base	first base	second base				third base
	U	C	A	G			U	C	A	G	
U	U Phe Leu Leu Leu	Ser Ser Ser Ser	Thr Thr Thr Thr	Val Val Val Val	U C A G	U Phe Leu Leu Leu	Ser Ser Ser Ser	Thr Thr Thr Thr	Val Val Val Val	U C A G	
C	Pro Pro Pro Pro	Pro Pro Pro Pro	Ile Ile Gln Gln	Arg Arg Arg Arg	U C A G	C Leu Leu Leu	Pro Pro Pro Pro	Ile Ile Gln Gln	Arg Arg Arg Arg	U C A G	
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asp Asp Leu Leu	Ser Ser Arg Arg	U C A G	A Ile Ile Met	Thr Thr Thr Thr	Asp Asp Leu Leu	Ser Ser Arg Arg	U C A G	
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Gly Gly	Gly Gly Gly Gly	U C A G	G Val Val Val Val	Ala Ala Ala Ala	Asp Asp Gln Gln	Gly Gly Gly Gly	U C A G	

Table 3: Action of fermion operators in the two schemes consistent with family box completeness and Watson-Crick duality. Bosonic families are shaded and fermionic families are unshaded.

Summarizing, we have shown that among the schemes exhibited in Table 1 and Table 2 above and except for the highly unbalanced and, in our view, unnatural solutions coming from items (c) and (d) of Table 2, the only ones that are invariant under some kind of natural duality transformation are the ones shown in Table 3 above, where the first is obtained from item (b<sub>1</sub>) of Table 1 using Watson-Crick duality with option (i) (without inversion of order) while the second is obtained from item (c) of Table 1 using Watson-Crick duality either with option (i) (without inversion of order) or with option (ii) (with inversion of order).

## 4 Conclusions

Combining the results of Section 2 and Section 3, we can state the main result of this paper as follows. Among the algebraic models for the evolution of the genetic code, the only ones that allow for an amino acid and codon assignment which is consistent with the requirements of family box completeness and invariance under some kind of natural duality transformation are the following.

1. The  $\mathfrak{sp}(6)$  model based on the chain

$$\mathfrak{sp}(6) \supset \mathfrak{sp}(4) \oplus \mathfrak{su}(2) \supset \mathfrak{su}(2)^3 \supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{su}(2) \supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{so}(2)$$

(see Table 4 of Ref. [7], p. 3189 for the branching rules)

- a) either with the central involution  $(1, -1, 1)$  in  $SU(2)^3$  and partial Watson-Crick duality defined without inversion of order  $((XYN)^\dagger = X^\dagger Y^\dagger N)$ , leading to the scheme shown in item (a) of Table 3,
- b) with the central involution  $(1, 1, -1)$  in  $SU(2)^3$  and partial Watson-Crick duality defined either without inversion of order  $((XYN)^\dagger = X^\dagger Y^\dagger N)$  or with inversion of order  $((XYN)^\dagger = Y^\dagger X^\dagger N)$ , leading to the scheme shown in item (b) of Table 3, which has already been identified in Ref. [5].

2. The  $\mathfrak{osp}(5|2)$  model based on the chain

$$\mathfrak{osp}(5|2) \supset \mathfrak{so}(5) \oplus \mathfrak{sp}(2) \supset \mathfrak{su}(2)^3 \supset \mathfrak{su}(2)^2 \supset \mathfrak{su}(2) \oplus \mathfrak{o}(2) \supset \mathfrak{o}(2) \oplus \mathfrak{o}(2)$$

(see Table V and Table VI of Ref. [9], pp. 5436 & 5438 for the branching rules) with the central involution  $(1, 1, -1)$  in  $SU(2)^3$ , which corresponds to the central involution  $(1, -1)$  in  $SU(2)^2$ , and partial Watson-Crick duality defined without inversion of order  $((XYN)^\dagger = X^\dagger Y^\dagger N)$ , leading to the scheme shown in item (a) of Table 3.

The resulting possible trees of evolution, with the pertinent amino acid assignments are shown in the following figures. Deciding whether they are also acceptable from the biological point of view remains a challenge for future investigations.



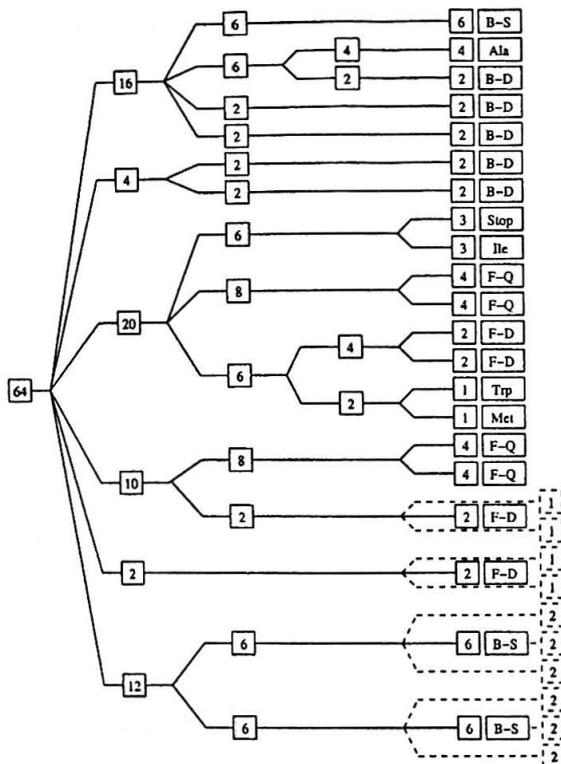


Figure 3: Possible tree of evolution with amino acid assignments for the standard genetic code in the  $sp(6)$  model, according to option 1, b) of the text and Table 3, (b): boxes marked "B-S" contain the three sextets Arg, Leu and Ser (all of which are bosonic), boxes marked "F-Q" contain the four fermionic quartets Gly, Pro, Thr and Val, boxes marked "F-D" contain the four fermionic doublets Cys, Gln, His and Tyr, and boxes marked "B-D" contain the five bosonic doublets Asn, Asp, Glu, Lys and Phe. Compare Fig. 1.

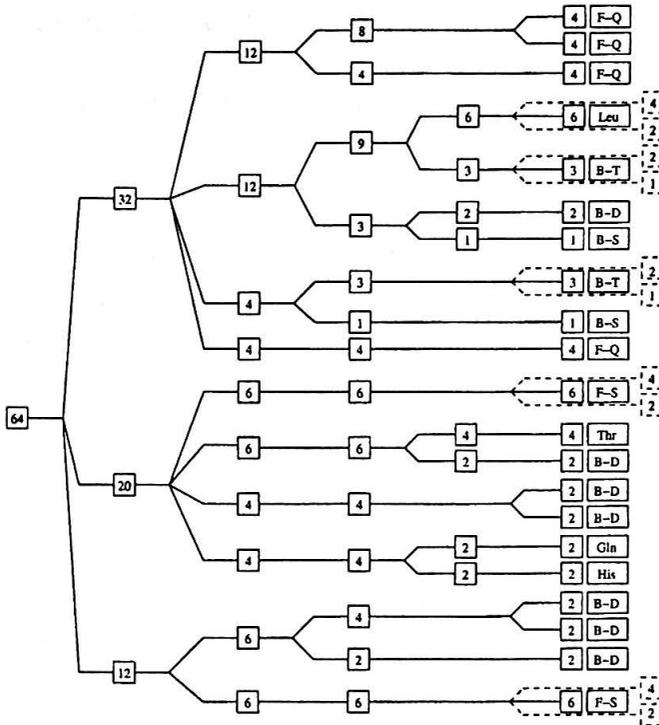


Figure 4: Possible tree of evolution with amino acid assignments for the standard genetic code in the  $osp(5|2)$  model, according to option 2 of the text and Table 3, (a): boxes marked “F-S” contain the two fermionic sextets Arg and Ser, boxes marked “F-Q” contain the four fermionic quartets Ala, Gly, Pro and Val, boxes marked “B-T” contain the two triplets Ile and Stop (both of which are bosonic), boxes marked “B-D” contain the seven bosonic doublets Asn, Asp, Cys, Glu, Lys, Phe and Tyr, and boxes marked “B-S” contain the two singlets Met and Trp (both of which are bosonic).

## References

- [1] E.N. Trifonov: *Consensus Temporal Order of Aminoacids and Evolution of the Triplet Code*, Gene **261** (2000) 139-151.
- [2] J.E.M. Hornos and Y.M.M. Hornos: *Algebraic Model for the Evolution of the Genetic Code*, Phys. Rev. Lett. **71** (1993) 4401-4404.
- [3] J. Maddox: *The Genetic Code by Numbers*, Nature **367** (1994) 111.
- [4] I. Stewart: *Broken Symmetry in the Genetic Code?*, New Scientist **141**, No. 1915 (1994) 16.
- [5] M. Forger, Y.M.M. Hornos and J.E.M. Hornos: *Global Aspects in the Algebraic Approach to the Genetic Code*, Phys. Rev. E **56** (1997) 7078-7082.
- [6] J.E.M. Hornos, Y.M.M. Hornos and M. Forger: *Symmetry and Symmetry Breaking: An Algebraic Approach to the Genetic Code*, Int. J. Mod. Phys. B **13** (1999) 2795-2885.
- [7] F. Antoneli, L. Braggion, M. Forger and J.E.M. Hornos: *Extending the Search for Symmetries in the Genetic Code*, Int. J. Mod. Phys. B **17** (2003) 3135-3204.
- [8] M. Forger and S. Sachse: *Lie Superalgebras and the Multiplet Structure of the Genetic Code I: Codon Representations*, J. Math. Phys. **41** (2000) 5407-5422.
- [9] M. Forger and S. Sachse: *Lie Superalgebras and the Multiplet Structure of the Genetic Code II: Branching Schemes*, J. Math. Phys. **41** (2000) 5423-5444.
- [10] F. Antoneli: *Grupos Finitos e Quebra de Simetria no Código Genético*, PhD Thesis, University of São Paulo, 2003.
- [11] F. Antoneli, M. Forger & J.E.M. Hornos: *The Search for Symmetries in the Genetic Code: Finite Groups*, Mod. Phys. Lett. B **18** (2004) 971-978.
- [12] F. Antoneli and M. Forger: *Finite Groups for the Genetic Code I: Codon Representations*, Preprint, University of São Paulo, 2000.
- [13] F. Antoneli and M. Forger: *Finite Groups for the Genetic Code II: Branching Rules*, Preprint, University of São Paulo, 2005.
- [14] The GAP Group: *GAP - Groups, Algorithms, and Programming 4.2*, Aachen, St. Andrews, 1999. (<http://www-gap.dcs.st-and.ac.uk/~gap>).
- [15] A. Crumeyrolle: *Orthogonal and Symplectic Clifford Algebras*, Kluwer, Dordrecht 1990.

RELATÓRIOS TÉCNICOS DO DEPARTAMENTO DE MATEMÁTICA APLICADA

2008

- RT-MAP-0801 - Freire Jr., R.S.; Garcia, M.V.P.; and Tal, F.A.  
"Instability of Equilibrium Points of Some  
Lagrangian Systems"  
Janeiro, 2008 - São Paulo - IME-USP - 16 pg.
- RT-MAP-0802 - Forger, Michael and Mário Otávio Salles  
"Hamiltonian Vector Fields on Multiphase Spaces of  
Classical Field Theory"  
Fevereiro, 2008 - São Paulo - IME-USP - 07 pg.
- RT-MAP-0803 - Angelo Barone Netto  
"Sobre Sudoku"  
Maio, 2008 - São Paulo - IME-USP - 06 pg.
- RT-MAP-0804 - Michael Forger, Fernando Antoneli Jr. and Paola Gaviria  
"Maximal Subgroups of Compact Lie Groups"  
Agosto, 2008 - São Paulo - IME-USP - x pg.
- RT-MAP-0805 - André de Carvalho and Toby Hall  
"Decoration Invariants for Horseshoe Braids"  
Agosto, 2008 - São Paulo - IME-USP - 35 pg.
- RT-MAP-0806 - André de Carvalho, Toby Hall and Rupert Venzke  
"On Periodic Minimal Pseudo-Anosov Braids"  
Agosto, 2008 - São Paulo - IME-USP - 05 pg.
- RT-MAP-0807 - Fernando Antoneli, Michael Forger, Paola Gaviria and  
José Eduardo M. Hornos  
"On Amino Acid and Colon Assignment in Algebraic  
Models for the Genetic Code"  
Agosto, 2008 - São Paulo - IME-USP - 27 pg.