Could Genetic Code Be Understood Number Theoretically?

M. Pitkänen,

February 2, 2024
Email: matpitka6@gmail.com.
http://tgdtheory.com/public_html/.
Postal address: Rinnekatu 2-4 A 8, 03620, Karkkila, Finland. ORCID: 0000-0002-8051-4364.

Contents

Abstract

The number of DNA triplets is 64. This inspires the idea that DNA sequence could be interpreted as an expansion of an integer using 64 as the base. Hence given DNA triplet would represent some integer in {0,1....,63} (sequences of I Ching symbols give a beautiful realization of these sequences).

The observation which puts bells ringing is that the number of primes smaller than 64 is 18. Together with 0, and 1 this makes 20: the number of amino-acids!

1. Questions

The finding just described stimulates a whole series of questions.

Do amino-acids correspond to integers in the set $S = \{primes < 64\} \cup \{0, 1\}$. Does amino-acid sequence have an interpretation as a representation as a sequence of integers consisting of 0, 1 and products of primes $p = 2, ..., 61$? Does the amino-acid representing 0 have an interpretation as kind of period separating from each other structural units analogous to genes representing integers in the sequence so that we would quite literally consists of sequences of integers? Do 0 and 1 have some special biological properties, say the property of being biologically inert both at the level of DNA and amino-acids?

Does genetic code mediate a map from integers 0,...,63 to set S such that 0 and 1 are mapped to 0 and 1? If so then three integers $2 \leq n \leq 63$ must correspond to stopping sign codons rather than primes. What stopping sign codon property means at the level of integers? How the map from integers 2,...,61 to the primes $p = 2, ..., 61$ is determined?

2. The chain of arguments leading to a number theoretical model for the genetic code

The following chain of arguments induced to large part by concrete numerical experimentation leads to a model providing a partial answer to many of these questions.

1. The partitions of any positive integer n can be interpreted in terms of number theoretical many boson states. The partitions for which a given integer appears at most once have interpretation in terms of fermion states. These states could be identified as bosonic and fermionic states of Super Virasoro representation with given conformal weight n.
2. The generalization of Shannon entropy by replacing logarithms of probabilities with the logarithms of p-adic norms of probabilities allows to have systems with negative entropy and thus positive negentropy. The natural requirement is that n corresponds to such prime $p \leq 61$ that the negentropy assigned to n is maximal in some number theoretic thermodynamics. The resulting correspondence $n \to p(n)$ naturally determined the genetic code.
3. One can assign to the bosonic and fermionic partitions a number theoretic thermodynamics defined by a Hamiltonian. Purely bosonic and fermionic thermodynamics are defined by corresponding partition functions Z_B and Z_F whereas supersymmetric option is defined by the product $Z_B \times Z_F$. Supersymmetric option turns out to be the most realistic one.
4. The simplest option is that Hamiltonian depends only on the number r of the integers in the partition. The dynamics would be in a well defined sense local and would not depend on the sizes of summands at all. The thermodynamical states would be degenerate with degeneracy factors given by total numbers $d_I(n, r)$ of partitions of type $I = B, F$. The invariants known as rank and crank define alternative candidates for the basic building blocks of Hamiltonian.
5. Ordinary exponential thermodynamics based on, say $e^{-H/T} = q_0^{r-1}$, q_0 a rational number, produces typically unrealistic genetic codes for which most integers are mapped to small primes $p \leq 11$ and many primes are not coded at all. The idea that realistic code could result at some critical temperature fails also.
6. Quantum criticality and fractality of TGD Universe inspire the idea that the criticality is an inherent property of Hamiltonian rather than only thermodynamical state. Hence Hamiltonian can depend only weakly on the character of the partition so that all partitions contribute with almost equal weights to the partition function. Fractality is achieved if Boltzmann factors are given by $e^{-H/T} = (r+r_0)^{n_0}$ so that $H(r) = log(r+r_0)$ serves as Hamiltonian and n_0 corresponds to the inverse temperature. The supersymmetric variant of this Hamiltonian yields the most realistic candidates for the genetic code and there are good hopes that a number theoretically small perturbation not changing the divisors $p \leq 61$ of partition function but affecting the probabilities could give correct degeneracies.

Numerical experimentation suggests however that this might not be the case and that simple analytic form of Hamiltonian is too much to hope for. A simple argument however shows that $e^{-H/T} = f(r)$ could be in quantum critical case be deduced from the genetic code by fixing the 62 values of $f(r)$ so that the desired 62 correspondences $n \to p(n)$ result. The idea about almost universality of the genetic code would be replaced with the idea that quantum criticality allows to engineer a genetic code maximizing the total negentropy associated with DNA triplet-amino-acid pair.

7. A natural guess is that the map of codons to integers is given as a small deformation of the map induced by the map of DNA codons to integers induced by the identification of nucleotides with 4-digits 0,1,2, 3 (this identification depends on whether first, second, or third nucleotide is in question). This map predicts approximate $p(n) = p(n+1)$ symmetry having also a number theoretical justification. One can deduce codon-integer and amino-acid-prime correspondences and at (at least) two Boltzmann weight distributions $f(n)$ consistent with the genetic code and Negentropy Maximization Principle (NMP) constrained by the degeneracies of the genetic code.

1 Introduction

I have developed several models for genetic code with motivation coming from the belief that there might be some deeper number theoretical structure involved. The model based on Combinatorial Hierarchy was discussed in the chapter "Genes and Memes". In this chapter two further models are developed. The chapter begins with a discussion of a model relying on exact A-G symmetry and almost exact T-C symmetry of the genetic code with respect to the third nucleotide. The idea is that genetic code has emerged in some sense as a product of 1-code and 2-code via symmetry breaking. This symmetry breaking is also a central element of both the second model discussed in this chapter and the number theoretic model developed in the next chapter. Second idea is that there is some kind of variational principle mathematically analogous to the second law of thermodynamics involved.

Unfortunately, the physical model developed for the pre-biotic evolution of the genetic code does not fully support the proposed symmetry breaking scenario. The 2-code in the physical model trivial in the sense that it is induced by RNA conjugation for RNA doublets whereas 1-code is deducible directly from wobble rules and is non-deterministic. Symmetry breaking of the physical model has a beautiful interpretation in terms of fundamental physics but the realization of the symmetry breaking is not quite what has been assumed in these three models and also in the model based on Combinatorial Hierarchy discussed in the chapter "Genes and Memes". Despite this the models deserve to be represented.

Since the number theoretic model is the basic topic of this chapter, it is perhaps in order to describe the basic observations leading to the model. The number of DNA triplets is 64. This inspires the idea that DNA sequence could be interpreted as an expansion of an integer using 64 as the base. Hence given DNA triplet would represent some integer in {0,1....,63} (sequences of I Ching symbols give a beautiful representation of numbers in 64 base).

The observation which puts bells ringing is that the number of primes smaller than 64 is 18. Together with 0, and 1 this makes 20: the number of amino-acids!

1.1 Questions

The finding just described stimulates a whole series of questions.

Do amino-acids correspond to integers in the set $S = \{primes < 64\} \cup \{0, 1\}$. Does amino-acid sequence have an interpretation as a representation as a sequence of integers consisting of 0, 1 and products of primes $p = 2, ..., 61$? Does the amino-acid representing 0 have an interpretation as kind of period separating from each other structural units analogous to genes representing integers in the sequence so that we would quite literally consists of sequences of integers? Do 0 and 1 have some special biological properties, say the property of being biologically inert both at the level of DNA and amino-acids?

Does genetic code mediate a map from integers 0,...,63 to set S such that 0 and 1 are mapped to 0 and 1? If so then three integers $2 \leq n \leq 63$ must correspond to stopping sign codons rather

than primes. What stopping sign codon property means at the level of integers? How the map from integers 2,...,61 to the primes $p = 2, ..., 61$ is determined?

1.2 The Chain Of Arguments Leading To A Number Theoretical Model For The Genetic Code

The following chain of arguments induced to large part by concrete numerical experimentation leads to a model providing a partial answer to many of these questions.

1. The partitions of any positive integer n can be interpreted in terms of number theoretical many boson states. The partitions for which a given integer appears at most once have interpretation in terms of fermion states. These states could be identified as bosonic and fermionic states of Super Virasoro representation with given conformal weight n or even better, with the states of conformal weight n created by U(1) Kac Moody generators so that basically a breaking of Kac Moody symmetry would be in question.

2. The generalization of Shannon entropy by replacing logarithms of probabilities with the logarithms of p-adic norms of probabilities allows to have systems with negative entropy and thus positive negentropy. The natural requirement is that n corresponds to such prime $p \leq 61$ that the negentropy assigned to n is maximal in some number theoretic thermodynamics. The resulting correspondence $n \to p(n)$ would naturally determine the genetic code.

3. One can assign to the bosonic and fermionic partitions a number theoretic thermodynamics defined by a Hamiltonian. Purely bosonic and fermionic thermodynamics are defined by corresponding partition functions Z_B and Z_F whereas supersymmetric option is defined by the product $Z_B \times Z_F$.

4. The simplest option is that Hamiltonian depends only on the number r of the integers in the partition. The dynamics would be in a well defined sense local and would not depend on the sizes of summands at all. The thermodynamical states would be degenerate with degeneracy factors given by total numbers $d_I(n, r)$ of partitions of type $I = B, F$. The invariants known as rank and crank define alternative candidates for basic building blocks of Hamiltonian.

5. Ordinary exponential thermodynamics based on, say $e^{-H/T} = q_0^{r-1}$, q_0 a rational number, produces typically unrealistic genetic codes for which most integers are mapped to small primes $p \leq 11$ and many primes are not coded at all. The idea that realistic code could result at some critical temperature fails also.

6. Quantum criticality and fractality of TGD Universe inspire the idea that the criticality is an inherent property of Hamiltonian rather than only thermodynamical state. Hence Hamiltonian can depend only weakly on the character of the partition so that all partitions contribute with almost equal weights to the partition function.

   Fractality is achieved if Boltzmann factors are given by $e^{-H/T} = (r + r_0)^{n_0}$ so that $H(r) = log(r + r_0)$ serves as Hamiltonian and n_0 corresponds to the inverse temperature. The supersymmetric variant of this Hamiltonian yields the most realistic candidates for the genetic code and one might hope that a number theoretically small perturbation not changing the divisors $p \leq 61$ of partition function but affecting the probabilities could give correct degeneracies.

   Numerical experimentation suggests however that this might not be the case and that simple analytic form of Hamiltonian is too much to hope for. A simple argument however shows that $e^{-H/T} = f(r)$ could be in quantum critical case be deduced from the genetic code by fixing the 62 values of $f(r)$ so that the desired 62 correspondences $n \to p(n)$ result. The idea about almost universality of the genetic code would be replaced with the idea that quantum criticality allows to engineer almost arbitrary genetic code. In this case the model becomes predictive if the condition that $S_{tot} = \sum_n S_{p(n)}(n)$ is minimized (negentropy maximization) with the constraint that each prime is coded and one could consider the possibility that $f(n)$ and $n \to p(n)$ is determined by this condition.

7. Genetic code has an almost unbroken symmetry in the sense that DNA triplets for which last nucleotide is A or G code for same amino-acid. For T and C this symmetry is slightly broken. This implies that the number of DNAs coding given amino-acid is almost always even. A very general number theoretic counterpart for this symmetry as a symmetry of partition function in the set 59 integers containing other than stopping codons. This symmetry must have fixed point and this is enough to explain why there is only single amino-acid coded by odd number DNAs besides singlets.

A natural guess is that the map of codons to integers is given as a small deformation of the map induced by the map of DNA codons to integers induced by the identification of nucleotides with 4-digits 0,1,2, 3 (this identification depends on whether first, second, or third nucleotide is in question). This map predicts approximate $p(n) = p(n+1)$ symmetry having also a number theoretical justification. One can deduce codon-integer and amino-acid-prime correspondences and at (at least) two Boltzmann weight distributions $f(n)$ consistent with the genetic code and Negentropy Maximization Principle constrained by the degeneracies of the genetic code.

1.3 What Is The Physical Counterpart Of The Number Theoretical Thermodynamics?

The partitions of any positive integer n can be interpreted in terms of number theoretical many boson states. The partitions for which a given integer appears at most once have interpretation in terms of fermion states. The states could be identified as bosonic and fermionic states of Super Virasoro representation with given conformal weight n or even better, with the states of conformal weight n created by U(1) Kac Moody generators so that basically a breaking of Kac Moody symmetry would be in question.

The obvious question concerns about the identification of the system in question. For instance, could it be associated with the light-like boundaries of magnetic flux quanta which are key actors in TGD based model of topological quantum computation [K1] ? If so, then each DNA triplet would correspond to a portion of magnetic flux quantum characterized by a conformal weight n determined by the DNA triplet in question. If there is single flux quantum parallel to the DNA strand, the value of n would be constant only along the portion of length corresponding to single DNA triplet. This non-conservation of conformal weight along light-like boundary is quite possible due to the breaking of strict classical non-determinism in TGD Universe having interpretation as a space-time correlate of quantum non-determinism.

With this identification one might perhaps interpret the integer determined by a given gene as a code for a topological quantum computer program using 64-base instead of 2-base. Since the boundaries of the magnetic flux tubes associated with DNA double strands are light-like, they can be interpreted either as states or as dynamical evolutions. Therefore the light-like boundary of the flux tube associated with DNA strand could be interpreted either as a code of a quantum computer program or as a running quantum computer program [K1].

The appendix of the book gives a summary about basic concepts of TGD with illustrations. There are concept maps about topics related to the contents of the chapter prepared using CMAP realized as html files. Links to all CMAP files can be found at `http://tgdtheory.fi/cmaphtml.html` [**?**]. Pdf representation of same files serving as a kind of glossary can be found at `http://tgdtheory.fi/tgdglossary.pdf` [**?**].

2 The First Model For The Evolution Of The Genetic Code

The exact A-G symmetry and almost exact T-C symmetry of the memetic codons with respect to third nucleotide suggest that genetic code factorizes in a good approximation to a product of codes associated with DNA doublets and singlets. This suggests factorization also at the level of pre-amino-acids. Perhaps DNAs triplets have resulted as a symbiosis of singlets and doublets whereas amino-acids might have been developed via a symbiosis of 2 molecules coded by 4 DNA singlets and 10 molecules coded by 16 DNA doublets.

In this section a formal model for the evolution of the genetic code based on the approximate factorization of the genetic code into a product code formed by doublet and singlet codes is discussed. Also physical model for the evolution of the genetic code is briefly discussed. Product

code as such predicts degeneracies approximately but fails at the level of detailed predictions for DNA-amino-acid correspondences. A "volume preserving" flow in discrete DNA space is needed to produce realistic DNA-amino-acid correspondences. This flow has the general tendency to cluster amino-acids to connected vertical stripes inside the 4-columns appearing as elements of the 4×4 code table, whose elements are labeled by the first two bases of DNA triplet. One can invent an information maximization principle providing a quantitative formulation for this tendency. The physical model for the evolution modifies the vision about RNA world [I2, I3].

2.1 Does Amino-Acid Structure Reflect The Product Structure Of The Code?

The exact A-G symmetry and the almost exact T-C symmetry of our genetic code supports approximate 2×10 structure such that 16 DNA doublets and 4 DNA singlets code for 10 *resp.* 2 "pre-amino-acids" which combine to form the real amino-acids. The 3×7 decomposition of the number 21 of amino-acids plus stopping sign suggests 3×7 decomposition of the genetic code. This decomposition is however not favored by the symmetries of the genetic code and will not be discussed in the sequel.

The coding of amino-acids involves tRNA binding with amino-acids and this means that the structure of amino-acids need not reflect the product structure of the genetic code and it might be that only the structure of tRNA reflects the product structure. The study of the amino-acid geometric structure does not reveal any obvious structural 3×7-ness or 2×10-ness. One can however wonder whether this kind of structures might be present at more abstract level and present only in the interactions of tRNA and amino-acids. As will be found, pre-amino-acids correspond most naturally to RNA sequences so that the product decomposition is realized trivially.

2.2 Number Theoretical Model For The Genetic Code

The study of the genetic code allows to deduce the process leading to the breaking of the product symmetry and T-C symmetry.

2.2.1 Approximate reduction to a product code

The dependence of the amino-acid coded by DNA on the third codon of DNA triplet is weak. This inspires the guess that triplet code might have evolved as a fusion of doublet code and singlet codes.

This should be reflected in its structure. The decomposition $20 = 2 \times 10$ for real amino-acids suggest that singlet code maps four bases to 2 "pre-amino-acids" such that A and G resp. T and C are mapped to same pre-amino-acid, and 16 doublets to 10 "pre-amino-acids". The exact A-G symmetry and almost exact T-C symmetry of our genetic code support this interpretation.

Product code hypothesis is very strong since the degeneracies of the product code are products of the degeneracies for the composite codes so that the number n_{AB} of DNA triplets coding a given amino-acid having the product form "AB", to be referred as the degeneracy of the amino-acid, is given by the product

$$n_{AB} = n_A \times n_B$$

of the degeneracies of the "pre-amino-acids" A and B. Here A and B can refer to $(A, B) = (3, 7)$ or $(A, B) = (2, 10)$ respectively.

The number $N_{AB}(n)$ of amino-acids with given degeneracy n is given by the formula

$$N_{12}(n) = \sum_{n_1 \times n_2 = n} N_1(n_1) N_2(n_2) \ ,$$

where $N_1(n_1)$ resp. $N_2(n_2)$ is the number of pre-amino-acids with the degeneracy n_1 resp. n_2.

For 2×10 case singlet sector allows only single candidate for the code since the genetic code has exact A-G symmetry and almost exact T-C symmetry with respect to the last base. Thus A and G code for the first pre-amino-acid and T and C the second one. A breaking of the T-C symmetry is needed to obtain realistic code.

n	1	2	3	4	6
N(prod)	0	12	0	4	4
N(real)	2	9	2	5	3

Table 1: The numbers $N(n)$ of amino-acids coded by n DNAs for unperturbed 2×10 product code and for the real genetic code for 2×10 option.

2.2.2 Our genetic code as result of symmetry breaking for 2×10 product code

As found, there are two cases to be considered: 3×7 T-C asymmetric and 2×10 T-C symmetric product code. The approximate T-C symmetry favors strongly 2×10 option and 3×7 will be considered only briefly in a separate subsection. On basis of degeneracies alone it is not possible to distinguish between these codes and 3×7 code was in fact the first guess for the product code.

In case of 2×10 code the decomposition of 16 DNA doublets giving almost the degeneracies of our genetic code is (3322 111 111).

$$(2 \oplus 2) \times (3 \oplus 3 \oplus 2 \oplus 2 \oplus 6 \times 1)$$

This gives

It is important to notice that the multiplets appear as doubled pairs corresponding to A-G and T-C symmetries. One generalized amino-acid (which cannot correspond to stopping sign) is lacking and must result by a symmetry breaking in which one amino-acid in the code table is transformed to a new one not existing there. Alternatively three amino-acids are transformed to stopping signs.

It is easy to find the deformation yielding correct degeneracies by removing DNAs from the DNA-boxes defined by various values of degeneracies to other boxes and adding them to other boxes. The rule is simple: taking m DNAs from a box containing n DNAs creates a box with $n - m$ DNAs and annihilates one n-box:

$$N(n) \to N(n) - 1 \ , \quad and \ N(n-m) \to N(n-m) + 1 \ .$$

If one adds k of these DNAs to r-box one has

$$N(r) \to N(r) - 1 \ , \ \ N(r+k) \to N(r+k) + 1 \ .$$

The operation which is not allowed is taking the entire content of a DNA box defined by amino-acid and adding it to other boxes since this would mean that the amino-acid in question would not be coded by any DNA. Thus the number of boxes can only grow in this process.

Realistic degeneracies are obtained by a rather simple operation.

1. Take from one 6-plet two amino-acid and move the first of them to 2-plet to get $N(6) = 3$, $N(4) = 5$, $N(3) = 1 < 2$, $N(2) = 11 > 9$ and move the second one to hitherto non-existing singlet to get $N(1) = 1$.

2. Move one DNA from some doublet to second doublet to get triplet and singlet to get $N(1) = 2$, $N(2) = 9$ and $N(3) = 2$.

This operation gives correct degeneracies only and it turns out that correct symmetry structure requires additional operations.

2.2.3 Failures of the product structure and the symmetry breaking as volume preserving flow in DNA space

A slightly broken product structure allows to understand the degeneracies of our genetic code relatively easily. It however leads also to wrong predictions at the level of DNA-amino-acid correspondence.

1. Exact product structure predicts that all 4-columns XYU, $U = A, G, T, C$ appearing as elements of the code table labelled by first and second bases of DNA triplet should have similar amino-acid structure. For 2×10 code the prediction is that all 4-columns should have $AABB$ structure and this prediction breaks down only for $AAAA$ type 4-columns.

2. For 2×10 code a given amino-acid should be coded either by DNA pairs of form (XYA, XYG) or of form (XYC, XYT). This is not the case. A given amino-acid tends to appear as connected vertical stripes inside the elements of the 4×4 table (4-columns). For instance, all 4-columns of form $AAAA$ (A=leu, val, ser, pro, thr, ala, arg, gly) and 3-column ile break the prediction of the product code.

3. In the case of 2×10 2n-plet formed by (XYA, XYG)-pairs is accompanied always by an 2n-plet formed by (XYT, XYC) pairs. By studying the degeneracies of the code one can get idea about how good these predictions are.

It seems that the breaking of the product symmetry tends to form connected vertical clusters of amino-acids inside a given element of the 4×4 code table but that one cannot regard stripes longer than 4 elements as connected structures. The 2×10 structure is favored by approximate T-C symmetry, and one can imagine that relatively simple flow in DNA space could yield the desired condensation of the amino-acids to form connected vertical stripes. The most general flow is just a permutation of DNAs and obviously preserves the degeneracies of various amino-acids. There are 64! different permutations but A-G and T-C symmetries reduce their number to 32!.

The idea about discrete volume preserving flow in DNA space can be made more precise. A-G and T-C gauge symmetries suggest the presence of a discrete symplectic structure. Perhaps one could regard 16×4 DNAs as 16 points of 4-dimensional discrete symplectic space so that the canonical symmetries of this space (volume preserving flows) acting now as permutations would be responsible for the exact A-G gauge invariance and approximate T-C gauge invariance. This brings in mind the canonical symmetries of CP_2 acting as $U(1)$ gauge transformations and acting as almost gauge symmetries of the Kähler action.

A natural guess is that the DNAs coding same amino-acid tend to be located at the same column of the 4×4 code table before the breaking of the product symmetry. If this is the case then only vertical flows need to be considered and A-G and T-C symmetries imply that their number is $8!^4$ corresponding to the four columns of the table.

Table 6c) summarizes our genetic code. It is convenient to denote the rows consisting of A-G resp. T-C doublets by X_1 and X_2. For instance, A_1 corresponds to the highest row phe-phe, ser-ser, tr-tyr, cys-cys and G_2 to the row leu-leu, pro-pro, gln-gln, arg-arg.

1. The simplest hypothesis is 2×10 option is realized and that the flow permutes entire rows of the code table consisting of A-G and T-C doublets. From **Table 2** it is clear that there is a G-C symmetry with respect to the first nucleotide broken only in the third row. This kind of primordial self-conjugacy symmetry would not be totally surprising since first and third nucleotides are in a somewhat similar position.

2. There are 3 6-plets leu, ser, and arg, and it is easy to see that one cannot transform them to the required form in which all 6-plets are on A-G or T-C row alone using this kind of transformation. For instance, one could require that leu doublets correspond to T-C doublets before the symmetry breaking. This is achieved by permuting the G_1 row with the C_2 row. Since A_2 contains also ser-doublet, also ser must correspond to T-C type 6-plet, and since arg is contained by G_2 row, also arg must correspond to T-C type 6-plet. Thus there would be 4 T-C type 6-plets but the product code gives only 2 of them.

3. The only manner to proceed is to allow mixing of suitable 6-plet of A-G type and 4-plet of T-C type in the sense that A-G doublet from 6 is moved to T-C doublet inside 4-plet and T-C doublet in 4-plet is moved to A-G doublet inside 6-plet. The exchange of AG_2 (ser doublet) and TG_1 (trh-doublet) represents this kind of permutation.

The tables below summarize the three stages of the construction.
At the last stage the T-C symmetry breaking giving rise to bla-trp and ile-met doublets occurs.

	A	G	T	C	
A	phe	ser	tyr	cys	A
	phe	ser	tyr	cys	G
	leu	thr	asn	thr	T
	leu	thr	asn	thr	C
G	val	ala	glu	gly	T
	val	ala	glu	gly	C
	leu	pro	gln	arg	T
	leu	pro	gln	arg	C
T	ile	ser	stop	ser	A
	ile	ser	stop	ser	G
	met	thr	lys	arg	T
	met	thr	lys	arg	C
C	val	ala	asp	gly	A
	val	ala	asp	gly	G
	leu	pro	his	arg	A
	leu	pro	his	arg	G

Table 2: Code table before the flow inducing the breaking of the product symmetry

	A	G	T	C	
A	phe	ser	tyr	cys	A
	phe	ser	tyr	cys	G
	leu	ser	stop	thr	T
	leu	ser	stop	thr	C
G	leu	pro	his	arg	A
	leu	pro	his	arg	G
	leu	pro	gln	arg	T
	leu	pro	gln	arg	C
T	ile	thr	asn	ser	A
	ile	thr	asn	ser	G
	met	thr	lys	arg	T
	met	thr	lys	arg	C
C	val	ala	asp	gly	A
	val	ala	asp	gly	G
	val	ala	glu	gly	T
	val	ala	glu	gly	C

Table 3: The code table after the action of the flow inducing the breaking of product symmetry

	A	G	T	C	
A	phe	ser	tyr	cys	A
	phe	ser	tyr	cys	G
	leu	ser	stop	stop	T
	leu	ser	stop	trp	C
G	leu	pro	his	arg	A
	leu	pro	his	arg	G
	leu	pro	gln	arg	T
	leu	pro	gln	arg	C
T	ile	thr	asn	ser	A
	ile	thr	asn	ser	G
	ile	thr	lys	arg	T
	met	thr	lys	arg	C
C	val	ala	asp	gly	A
	val	ala	asp	gly	G
	val	ala	glu	gly	T
	val	ala	glu	gly	C

Table 4: The code table after the T-C symmetry breaking

1. thr 6-plet is transformed to 4-plet by replacing thr-thr in AC_2 by bla-trp. trp is the missing amino-acid.

2. TA_2 met-doublet is transformed to ile-met so that the realistic genetic code results.

One might argue that symmetry breaking permutations $G_1 - C_2$ and $AG_2 - TG_1$ should permute amino-acids with a similar chemical character. A similar constraint applies to T-C symmetry breaking. By studying the chemical structure of the amino-acids, one finds that this is satisfied to a high degree.

1. The permutations val-leu and ala-pro exchange amino-acids with non-polar (hydrophobic) side groups. The permutations glu-his and gly-arg exchange polar (hydrophilic) amino-acid with a polar amino-acid which is also basic. Ser and thr are both non-polar amino-acids.

2. Ile and met are both non-polar so that ile$\rightarrow$ met replacement satisfies the condition.

3. The objection is that the side group for trp is non-polar but polar for thr. Interestingly, the code table decomposes to two connected regions corresponding to non-polar/polar side groups at the left/right such that the non-polar trp located inside the polar region is the only black sheep whereas thr naturally belongs to the polar region (see **Fig. ??**). As will be found trp is also otherwise singular case.

2.2.4 The information maximization principle determining the "volume preserving flow"

The interaction between the DNA singlets and doublets is the physical explanation for the breaking of the product symmetry. This interaction involves two parts: the flow and T-C symmetry breaking. The flow is analogous to the formation of connected vertical stripes of amino-acids in DNA space: kind of condensation process in which different phases represented by amino-acids tend to condense to form regions consisting of at most 4-units of type XYU, $U = A, G, T, C$. Obviously this means continuity and thus also symmetry analogous to that emerging when (amino-acid) gases condense to a liquid state: the breaking of the product symmetry is the price paid for this additional symmetry. It turns out to be possible to formulate a variational principle consistent with the proposed flow in the direction of the columns of the code table and defining the dynamics of the condensation.

What this means that one can assign an information measure to the code table such that the volume preserving flow in question maximizes this information measure.

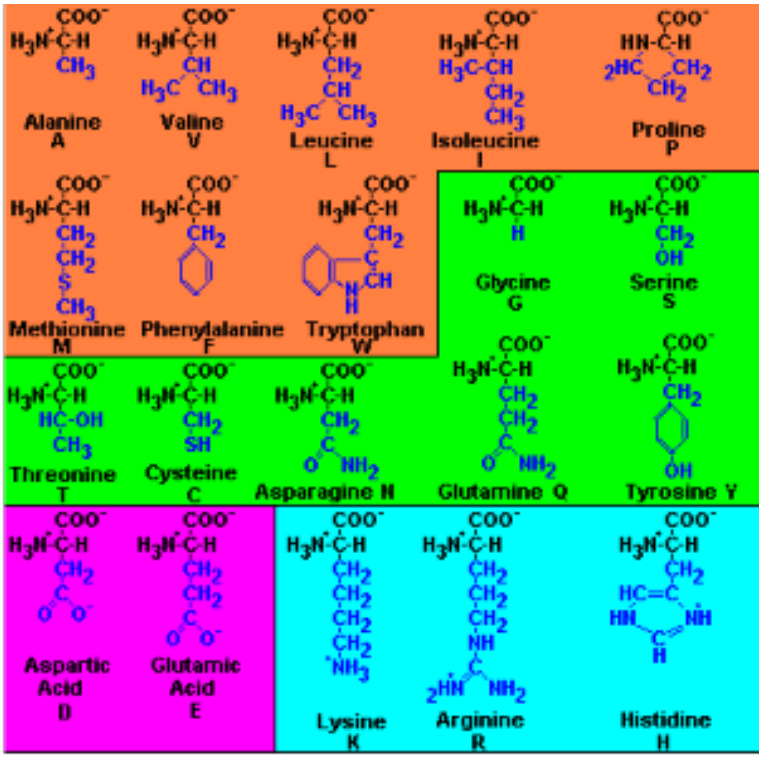

Figure 1: The chemical structure of amino-acids. The first group (ala, ...) corresponds to non-polar amino-acid side groups, the remaining amino-acids to polar side groups. The two lowest groups correspond to acidic (asp, glu) and basic side groups.

1. Information measure is assumed to be local in the sense that it decomposes into a sum of information measures associated with the elements C_{AB}, $A, B \in \{A, G, T, C\}$, of the 4×4 code table (elements are 4-element columns). In the physical analogy this means that the condensed droplets of various amino-acids can have at most the size of single 4-element column.

2. Consider the element C_{AB}. Let the multiplet associated with the amino-acid a_k contain $n(k, AB)$ amino-acids and let $i(k, AB)$ tell the number of the disjoint parts to which the amino-acids a_k in the 4-plet AB split. The number of these disjoint multiplets can be $0, 1, 2$. Let the i: th region contain $n(k, AB, i)$ amino-acids a_k. The meaning of the equations

$$\sum_{i=1}^{i(k,AB)} n(a_k, AB, i) = n_k(AB) \quad ,$$

$$\sum_{AB} n_k(AB) = n_k \quad ,$$

$$\sum_k n_k = 64$$

is obvious.

Assign to the i: th connected region containing $n(k, i, AB)$ identical amino-acids a_k probability

$$p(k, i, AB) = \frac{n(k, i, AB)}{64} \quad ,$$

to the element AB the total probability

$$p(k,AB) = \sum_{i=1}^{i(k,A,B)} p(k,i,AB) \ ,$$

and to the entire table the probability

$$p_k = \sum_{AB} p(k,AB) = \frac{n(k,AB)}{64} \ .$$

The sum of the probabilities associated with various amino-acids satisfies

$$\sum_k p_k = 1 \ .$$

The information measure associated with amino-acid a_k element AB is defined as

$$I(k,AB) = \sum_{i=1}^{i(k,A,B)} p(k,i,AB) \times log[p(k,i,AB)] \ ,$$

Note that this number is non-positive always. The total information associated with the amino-acid a_k in code table is defined as

$$I(k) = \sum_{AB} I(k,AB) \ .$$

The total information of the code table is defined as the sum of the information measures associated with various amino-acids:

$$I = \sum_k I(k) \ .$$

This information measure is maximized (which means the minimization of the absolute value of the measure since one can speak of the minimization of entropy) by the vertical flow satisfying the previous constraints, and thus satisfying the constraints that the numbers a_k of various amino-acids are fixed and $A \leftrightarrow G$ and $T \leftrightarrow C$ symmetries are respected. There is a direct analogy with thermodynamical equilibrium with fixed particle numbers and symmetry. The equilibrium is characterized by the chemical potentials associated with the amino-acids. There is no temperature type parameter now.

The variational principle indeed favors the formation of vertically connected regions consisting of $n = 2, 3$ or 4 amino-acids. By construction the variational principle does not tell anything about larger regions. In particular, it is more favorable for 4 amino-acids in a given column (say ser in the second column of the table) to be contained by single element than by 2 elements since the information measure would be $-1/16log(1/16)$ for two disjoint doublets and $-1/16log(1/8)$ for singlet 4-plet in same element and thus smaller in absolute value. In the similar manner the AAAB decomposition of singlet element instead of say AABA is favored.

2.2.5 The deviations from the standard code as tests for the basic symmetries of the model

The deviations of the genetic codes from the standard code [I1] provide a testing ground for the postulated symmetries of the genetic code and might also help to deduce the alien codes.

The deviations from universality of the Start codon (coding for met) and stop codons are very rare. With two exceptions all known deviations from the standard code are located in the first and fourth columns of the code table. For the first exceptional case the codon is ATC in the third column and codes for both stopping sign and pyrrolysine, which is an exotic amino-acid. It is somewhat a matter of taste whether one should say that the universality of the third column is broken or not since, depending on context, ATC codes stopping sign or pyrrolysine. Second exceptional case corresponds to the use of two stop codons to code amino-acids and this necessarily breaks the universality of the third column in T-C 2-subcolumns. No violations of the predicted A-G symmetry and the universality of the second column of the code table are known.

The deviations from the standard code [I1] provide valuable hints when one tries to deduce information about the alien codes.

1. Consider first the mitochondrial genes.
   i) Mitochondrial codon ACT from animals and micro–organisms (but not from plants) codes trp instead of stopping sign.
   ii) Most animal mitochondria use TAT to code met instead of ile.
   iii) Yeast mitochondria use GAX codons to code for thr instead of leu.

2. The violations of the universality are very rare for nuclear genes. A few unicellular eukaryotes have been found that use one or two of three stop codons to code amino-acids instead. The use of two stop codons to code amino-acids necessarily violates the universality of the third column but need not break the universality for the embedding of amino-acid space to DNA space.

3. There are also two non-standard amino-acids: selenocysteine and pyrrolysine.

   (a) Selenocysteine is encoded by ACT (fourth column) coding stopping sign normally. Interestingly, ACT codes also stopping sign and the translation machinery is somehow able to discriminate when selenocysteine is coded instead of stop. This codon usage has been found in certain Archaea, eubacteria, and animals. This deviation means that the number of amino-acids is 21 or 20 depending on context. This conforms with the view that number 21 indeed has a deep number theoretical meaning and that one can regard stopping sign formally as amino-acid.

   (b) In one gene found in a member of the Archaea, exotic amino-acid pyrrolysine is coded by ATC, which corresponds to the lower stopping sign in the code table. This case represents the only deviation from universality of the third column of the code table but even in this case also stopping sign is coded. How the translation machinery knows whether to code pyrrolysine or to stop translation is not yet known. TGD would suggest that electromagnetic signalling mechanisms ("topological light rays") might be involved.

3 Basic Ideas And Concepts Underlying Second Model Of Genetic Code

In the following the basic ideas and concepts are summarized.

3.1 Genetic Code From The Maximization Of Number Theoretic Information?

One of the earlier ideas about genetic code was that genetic code maximizes some kind of information measure [**?**, K3, K4]. In that context ordinary entropy was used. The discovery of number theoretic variants of Shannon entropy based on p-adic norm allows however a modified approach.

3.2 Genetic Code From A Minimization Of A Number Theoretic Shannon Entropy

The idea about entropy minimization determining genetic generalizes to the idea that the map $n \to p(n)$ from integers representing DNA to primes representing amino-acids maximizes some kind of information measure.

3.2.1 Identification of ensembles

There is a natural candidate for the ensemble. This ensemble is defined by the partitions of n to sums of integers identified in terms of many-boson states. Each partition of an integer would correspond to a physical state. For Virasoro representations encountered in conformal field theories this is indeed the case. One can also consider partitions subject to some additional conditions.

For instance, one could require that same integer appears at most once or that only odd integers appear in the partition (these options are in fact equivalent).

These two ensembles correspond to bosonic and fermionic systems and states in question correspond to the bosonic and fermionic states of given conformal weight n in Super Virasoro reprsentation. Supersymmetric alternative would be based on the product of bosonic and fermionic partitition functions so that entropy would be the sum of the bosonic and fermionic contributions. In the sequel all these options will be studied and supersymmetric option turns out to be the most promising one.

In the bosonic case the partition numbers are conveniently calculated by using the recurrence relation [A1]

$$d_B(n,r) = P(n,r) = P(n-1,r-1) + P(n-r,r) \ , \ P(n,1) = 1 \ . \tag{3.1}$$

In the fermionic case the numbers $Q(n,k)$ of partitions of n to a sum of integers such that same integer does not appear twice characterize simplest models. These numbers are obtained from the formula [A1]

$$d_F(n,r) = Q(n,r) = P(n - \left(\begin{array}{c} r \\ 2 \end{array} \right), r) \quad . \tag{3.2}$$

These formulas allow a highly effective numerical treatment when Boltzmann weights depend on r only.

3.2.2 Identification of information measures

There is also a good guess for the information measure as the p-adic entropy S_p obtained by replacing the argument logarithm of a rational valued probability p_k appearing in Shannon entropy with the logarithm of its p-adic norm $|p_k|_p$. If the probabilities of partitions are same and given by $1/d_I(n)$, $I = B, F$, where $d_I(n)$ is the total number of partitions, one would have

$$S_{I,p}(n) = -\sum_1^{d_I(n)} \frac{1}{d_I(n)} log(|\frac{1}{d_I(n)}|_p) = -log(|\frac{1}{d_I(n)}|_p) \ , \ I = B, F \ . \tag{3.3}$$

The simplest model obviously corresponds to a high temperature limit in thermodynamics. $S_{I,p}(n)$ can be expressed also in a form which is a convenient starting point for finite temperature thermodynamics with Hamiltonian given by the number r of integers in the partition.

$$\begin{aligned} S_{I,p}(n) &= -\sum_{r=1}^{n} p_I(n,r) log(|\frac{1}{d_I(n)}|_p) \ , \\ p_I(n,r) &= \frac{d_I(n,r)}{d_I(n)} \ , \ d_I(n) = \sum_{r=1}^{n} d_I(n,r) \ , \ I = B, F \ . \end{aligned} \tag{3.4}$$

$p_I(n,r)$ is the total probability that partition has r summands.

What makes number theoretical thermodynamics so fascinating is that p-adic entropies can be negative so that they can become genuine information measures. Indeed, if $d_I(n)$ is divisible by p the p-adic norm of $d_I(n)$ can become smaller than one and its contribution to the entropy is negative. Hence the maximization of $S_{I,p}$ as a function of p assigning to n a unique prime $p(n)$ is natural in the case of genetic code. Furthermore, if $S_{I,p}(n)$ zero or positive, n does not carry information and is an excellent candidate for the stopping sign codon.

It is possible to deduce the correspondence $n \to p(n)$ by using simple number theoretical arguments. If the number $d_I(n)$ of partitions is divisible by p, n might be mapped to p since the logarithm of $1/d_I(n)$ receives a large negative contribution tending to make the number theoretic entropy negative. It is easy to see that the largest power of prime appearing in $d_I(n)$ determines $p(n)$ in the case that $d_I(n)$ is divisible by some primes $p \leq 61$. At high temperature limit any prime $p \leq 61$ yields the same value of $S_{I,p}(n)$.

3.3 High Temperature Limit For Bosonic, Fermionic, And Supersymmetric Thermodynamics

The tables below represent the bosonic and fermionic partition numbers and the prediction of high temperature limit of number theoretical thermodynamics in the bosonic, fermionic, and supersymmetric cases.

High temperature limit does not predict a realistic genetic code.

1. The decompositions of $d(n)$ to primes contain all primes < 64 except 37 and 61. 23 is not allowed by the rule determining the value $p(n)$. In fermionic case $d(n)$ is divisible by 61 for $n = 24$ and by 37 for $n = 28, 20, 47, 62$.

2. For $n = 13$ and $n = 36$ for which $d(n)$ is prime larger than 61 so that it is not possible to assign any unique prime to them ($n = 13$ seems to deserve its bad reputation!), p-adic entropy and thus also information vanishes. A possible interpretation is that these two zero information integers correspond to stopping sign codons. In the general case integers coding $p = 2$ are good candidates for stopping codons since minimization of entropy favors $p = 2$ when the partition function fails to be divisible by any prime $p \leq 61$.

3. The primes p smaller than 13, in particular $p = 11$, which would be coded by as many as 19 DNAs, are strongly over-represented. The over-representation of small integers might reflect the three congruences $p(4+5d) \mod 5 = 0$, $d(5+7r) \mod 7 = 0$, and $d(6+11r) \mod 11 = 0$ found by Ramanujan for which quite recently a proof and generalization has been found [A3].

4. For both fermionic and supersymmetric partition functions primes 41 and 43 fail to be coded and there is strong over-abundance of $p = 2$. An amusing numerical coincidence is that $d_F(20) = 64$ holds true.

n	$d_B(n)$	$p_B(n)$	$d_F(n)$	$p_F(n)$	$p_{BF}(n)$
0	1	1	1	0	0
1	1	1	1	1	1
2	2	2	1	1	2
3	3	3	2	2	3
4	5	5	2	2	5
5	7	7	3	3	7
6	11	11	4	2	11
7	3×5	5	5	5	5
8	2×11	11	6	3	11
9	$2 \times 3 \times 5$	5	8	2	2
10	$2 \times 3 \times 7$	7	10	5	3
11	$2^3 \times 7$	2	12	2	2
12	7×11	11	15	5	11
13	101 (prime)	?	18	3	3
14	$3^3 \times 5$	3	22	11	3
15	$2^4 \times 11$	2	27	3	3
16	$3 \times 7 \times 11$	11	32	2	2
17	$3^3 \times 11$	3	38	19	3
18	$5 \times 7 \times 11$	11	46	23	23
19	$2 \times 5 \times 7^2$	7	54	3	7
20	$3 \times 11 \times 19$	19	64	2	2
21	$2^3 \times 3^2 \times 11$	11	76	2	2
22	$2 \times 3 \times 167$	3	89(prime)	?	3
23	5×251	5	104	13	13
24	$3^2 \times 5^2 \times 7$	5	122	61	61
25	$2 \times 11 \times 89$	11	142	11	11
26	$2^2 \times 3 \times 7 \times 29$	29	165	29	29
27	$2 \times 5 \times 7 \times 43$	43	192	2	2
28	$2 \times 11 \times 13^2$	13	222	37	13
29	$5 \times 11 \times 83$	11	256	2	2

Table 5: Table represents the partition numbers $d_B(n)$ and $d_F(n)$ as well as the primes $p_B(n)$, $p_F(n)$, $p_{BF}(n)$ resulting from the minimization of the p-adic entropy $S_{I,p}(n)$, $I = B, F, BF$ as a function of n for $n < 30$.

4 Could Finite Temperature Number Theoretic Thermodynamics Reproduce The Genetic Code?

The number theoretical ansatz in its simplest form fails. It is however possible to modify the measure associated with the partitions, which can be regarded as $T \to \infty$ limit of thermodynamics. Some kind of conserved quantity playing the role of Hamiltonian and distinguishing between different partitions should be introduced.

p-Adic thermodynamics implies that the counterpart of Boltzmann exponent $exp(-H/T)$ should be rational. One manner to guarantee this is to assume Boltzmann weight has the form q_0^{-H/T_r} for some rational number q_0 assuming that H/T_r is integer valued. A stronger condition is that q_0 is integer. With natural conventions both Hamiltonian and the inverse of the reduced temperature T_r are integer valued. For $T_r = 1/k$ the counterpart of the ordinary temperature would be $T = k/log(q_0)$. Thus q_0 would partially characterize the number theoretical temperature of DNA-amino-acid system and varying temperature would allow the possibility of several codes. The genetic code indeed involves small variations [**?**, K4]. The Hamiltonian should depend on the number r of integers in the partition and possibly n, perhaps also on more refined properties of the partition.

The finite temperature need not as such be enough to guarantee a reasonable genetic code. On purely statistical grounds one expects that small primes appear very frequently as divisors of integer valued reduced partition function and over- abundance of small primes is expected. Detailed calculations in low temperature phase confirm this prediction.

Physical intuition suggests that there could exists something analogous to a critical temperature in the sense that large long range fluctuations for ordinary criticality correspond to large degeneracies for large primes. The challenge would be to find this critical phase expected to be located somewhere between high temperature phase and low temperature phases with $(r_0 > 1, s_0 = 1)$ and thus characterized by $r_0 > s_0 > 1$. The attempts to realize this program have not led to a success, and it seems that it is not only particular thermodynamical state of the system which should be critical, but the very Hamiltonian defining the number theoretical thermodynamics as the quantum criticality of TGD Universe indeed suggests.

4.1 How To Choose The Hamiltonian?

4.1.1 Hamiltonian as a function of the number of summands in the partition?

The most symmetric positive definite Hamiltonian one can imagine is $H(n,r) = H(r) = r$ thermodynamically equivalent with $H(r) = r - 1$. The independence of the Hamiltonian on n conforms with the idea that the dynamics is local in the sense that only the number r of integers in the partition matters and that the value n of the individual integer is irrelevant. Dynamics would be same for all values of n and in this sense universal.

A possible interpretation for $H(r)$ is in terms of the breaking of conformal symmetry allowing to distinguish between states characterized by the same eigenvalue n of the Virasoro generator L_0 and generated by the products $\prod_k L_{n_k}$ of Virasoro generators. This Hamiltonian is certainly the most natural starting point because it possesses maximal symmetries and is also computationally tractable.

For the corresponding thermodynamics temperature corresponds to a rational $q = r/s > 1$ and Boltzmann weights are given by the exponents q^r. It turns out difficult to find realistic looking genetic codes for this thermodynamics. Unless q is near unity only lowest values of r contribute and the general tendency is that the spectral power concentrates at small primes ≤ 11. This can be understood from the fact that small primes are the most probable divisors of random integers. The only hope seems to be that there exists a critical temperature at which large long range fluctuations correspond to large degeneracies for large primes.

The most general thermodynamics allows arbitrary function $exp(H/T) = f(r,T)$ of r having positive integers as values. An especially natural choice is $f(r) = (r + r_0)^n$ corresponding to Hamiltonian $H = log(r + r_0)$ and temperature $T = 1/n$ so that one has

$$exp(-H/T) \quad = \quad (r + r_0)^{n_0} \quad , \quad n_0 = \pm 1, \pm 2, ... \quad , \quad r_0 = 0, 1, 2, ... \quad . \tag{4.1}$$

n	$d_B(n)$	$p_B(n)$	$d_F(n)$	$p_F(n)$	$p_{BF}(n)$
30	$2^2 \times 3 \times 467$	2	296	2	37
31	$2 \times 11 \times 311$	11	340	17	17
32	$3 \times 11^2 \times 23$	11	390	13	11
33	$3^2 \times 7^2 \times 23$	7	448	2	7
34	$2 \times 5 \times 1231$	3	512	2	2
35	$3 \times 11^2 \times 41$	11	585	13	11
36	17977(prime(	?	668	2	2
37	$7 \times 11 \times 281$	11	760	19	19
38	$5 \times 11^2 \times 43$	11	864	2	11
39	$3^4 \times 5 \times 7 \times 11$	3	982	2	3
40	$2 \times 3 \times 7^2 \times 127$	7	1113	53	7
41	$3 \times 7 \times 11 \times 193$	11	1260	3	7
42	$2 \times 11 \times 2417$	11	1426	31	31
43	$3^4 \times 11 \times 71$	3	1610	23	7
44	$5^2 \times 31 \times 97$	31	1816	2	31
45	$2 \times 41 \times 1087$	41	2048	2	2
46	$2 \times 3 \times 73 \times 241$	3	2304	2	2
47	$2 \times 7^2 \times 19 \times 67$	7	2590	37	7
48	$3 \times 7 \times 7013$	7	2910	5	3
49	$5^2 \times 11 \times 631$	5	3264	2	2
50	$2 \times 11 \times 9283$	11	3658	59	59
51	$3 \times 11^2 \times 661$	11	4097	17	11
52	$3 \times 7 \times 11 \times 23 \times 53$	53	4582	29	53
53	$3^2 \times 7 \times 5237$	3	5120	2	2
54	$5 \times 7 \times 11 \times 17 \times 59$	59	5718	3	59
55	$2^2 \times 7 \times 71 \times 227$	7	6378	3	2
56	$11 \times 47 \times 1019$	47	7108	47	47
57	$2 \times 3 \times 102359$	3	7917	29	29
58	$2^2 \times 5 \times 11 \times 3251$	11	8808	2	2
59	$2^2 \times 5 \times 11 \times 19 \times 199$	19	9792	2	2
60	$17 \times 139 \times 409$	17	10880	2	17
61	$3 \times 5 \times 7 \times 11 \times 971$	11	12076	2	11
62	$2^2 \times 11 \times 13 \times 2273$	13	13394	37	37
63	$3 \times 113 \times 4441$	3	14848	2	2
64	$2 \times 5 \times 11 \times 71 \times 223$	11	16444	11	11
65	2×1006279	2	18200	5	5

Table 6: Table represents the partition numbers $d_B(n)$ and $d_F(n)$ as well as the primes $p_B(n)$, $p_F(n)$, $p_{BF}(n)$ resulting from the minimization of the p-adic entropy $S_{I,p}(n)$, $I = B, F, BF$ as a function of n for $30 \leq n \leq 65$. Note that for bosonic case $p = 37$ and 61 are not coded whereas for supersymmetric case $p = 41$ and 43 are not coded.

n	1	2	3	4	6
N	2	9	2	5	3

Table 7: The numbers $N(n)$ of amino-acids coded by n DNAs.

so that a second integer valued parameter creeps in. Note that the thermodynamics is invariant under the scalings $(r + r_0) \to \lambda \times (r + r_0)$.

For $n \geq 0$ the formula for $S_p(n)$ is computationally very attractive but $n_0 > 0$ corresponds to negative temperatures or negative values of $H(r)$. It is of course not clear whether the sign of temperature is really important since the number of states is finite. The general vision that rational valued entanglement coefficients correspond to negative entropy and are associated with bound states would suggest that H has interpretation as the analog of negative of binding energy and is therefore negative.

For $n_0 < 0$ the numerical calculations are somewhat intricate due to the emergence of factorials up to 63! in the calculation of p-adic norms of partition coefficients. The factors $1/(r + r_0)^{n_0}$ tend to divide from the partition function prime factors $r_0 + 1$ away and this means that for small values of r_0 the primes $pr_0 + 1 \leq 61$ rarely divide it. Hence an entropic phase is in question for $r_0 < 61$ and numerical calculations demonstrate that only few $p > 2$ are coded. One might hope that the situation changes for $r_0 > 61$ and should resemble that for $n_0 > 0$. Numerical calculations show that this is not the case. The outcome is a complete spontaneous magnetization in the sense that only $p = 2$ is coded. This can be understood from the fact that entropy is minimum for $p = 2$. The safe conclusion seems to be that $n_0 > 0$ phase is the only option possibly reproducing the genetic code for a properly chosen Hamiltonian.

The polynomial rather than exponential thermodynamics would conform with the quantum criticality and fractality of TGD Universe. The nice feature of the logarithmic Hamiltonian is that it describes inherently critical system since the thermodynamical weights are slowly varying functions of r and therefore thermal fluctuations are large. Therefore there are hopes of achieving criticality, perhaps for all values of n for $n_0 > 0$.

These optimistic expectations turn out to be correct. Numerical calculations for $n_0 > 0$ bosonic case demonstrate that the concentration of spectrum to small primes is not anymore present, all primes can be coded in some cases, and qualitatively reasonable looking genetic codes are obtained with degeneracies smaller than 8. The next improvement is super-symmetry which leads to more realistic candidates for genetic code with small parameter values. It is quite possible that the requirement that the realistic genetic code results exactly fixes the Hamiltonian completely and that some kind of symmetry breaking is required to get the correct code.

4.1.2 Hamiltonian as a function of the rank of the partition?

There are also more complex candidates for the Hamiltonian if one allows Hamiltonian to have different values for partitions having the same value of r. Already Dyson introduced the notion of rank of a partition of type (n, r) as the difference

$$R(n, r, n_{max}) \quad = \quad n_{max} - r \ , \tag{4.2}$$

where n_{max} is the largest integer appearing in the partition [A3].

Rank divides the partitions into equal sized classes and the number of them obviously appears as a factor in $d(n)$. The notion of rank allows to prove the congruences $d(4 + 5d) \ mod \ 5 = 0$ and $d(5 + 7r) \ mod \ 7 = 0$ discovered by Ramanujan but fails for $d(6 + 11r) \ mod \ 11 = 0$ as found by Dyson [A2]. Dyson speculated the existence of a more complex invariant which he christened crank.

Rank is not positive definite as the study of simplest examples demonstrates. A simple manner to get a non-negative Hamiltonian is based on the so called group number defined as rank modulo $n + 1$:

$$G(n, r, n_{max}) \quad = \quad R(n, r, n_{max}) \ mod \ n + 1 \ , \tag{4.3}$$

and having values only in the set $\{0, ..., n\}$. The modulo arithmetics has the effect of producing double degeneracy of partitions with same group number. The numbers $N(p)$ coding given prime satisfy $N(p) \geq 2$ for the real genetic code and this might be due to the modular arithmetics (the exponential thermodynamics based on rank predicts typically $N(p) = 1$ or 0 for $p > 11$).

Hence one could consider the Boltzmann weights

$$\begin{aligned} exp(-H/T) &= q^{kH(n,r,n_{max})} \ , \\ kH(n,r,n_{max}) &= G(n,r,n_{max}) \ . \end{aligned} \tag{4.4}$$

For this option partitions $(r, n_2...n_r)$ are favored for positive temperatures since $R = 0$ in this case and at low temperature limit the finding of genetic code reduces to the identification of the largest prime power factors of the number of partitions of n of type $(r, n_2...n_r)$. Note that ground state degeneracy results whereas for $H = r$ the ground state is singly degenerate at low temperature limit. The study of small values of n shows that this thermodynamics is not very interesting since the number of partitions of this kind is rather small. For large primes this would mean that they cannot be coded.

The inherently critical option corresponds to

$$exp(-H/T) = (G(n,r,n_{max}) + g_0)^k \ , \tag{4.5}$$

with integer valued temperature k. For $g_0 = 0$ the partitions of type $(r, n_2...n_r)$ would have zero thermodynamical weights for $k > 0$ and infinite conformal weight for $k < 0$.

4.1.3 Hamiltonian as the function of the crank of the partition?

Quite recently Mahlburg [A3] represented an ingenious proof of a theorem generalizing the famous regularities of partitions discovered by Ramanujan and followers (for a popular representation of what is involved see the article [A2]). The proof is based on the identification of the invariant anticipated by Dyson.

The reason why a function of crank is a promising candidate for Hamiltonian is following.

A partial explanation for why primes $p \leq 11$ are so abundant at infinite temperature limit is that $d(n)$ is divisible by $5, 7, 11$ for $n = 4 + 5k$, $n = 5 + 7k$, and $n = 6 + k11$ respectively so that these primes are strong competitors in negentropy maximization race for a large number of values of n (19 for $p = 5$, 8 for $p = 7$, 5 for $p = 11$).

Crank, denote it by C, decomposes the partitions to subsets for which numbers of elements are divisible by $5, 7$ *resp.* 11 in these three cases. The expression for $S_p(n)$ in the case of polynomial thermodynamics can be written as

$$\begin{aligned} S_p(n) &= \sum_i N(n,i) C^k(i) log(|\frac{C^k(i)}{Z(n)}|_p) \ , \\ Z(n) &= \sum_i N(n,i) C^k(i) \ , \end{aligned} \tag{4.6}$$

It is clear that 5, 7, 11 appearing as divisors in both $N(n,i)$ and $Z(n)$ cancel each other and there is no large contribution to negentropy from these primes. This contribution is actually tamed also for other Hamiltonians defining polynomial dynamics.

4.2 Could Supersymmetric $N_0 > 0$ Polynomial Thermodynamics Determine The Genetic Code?

The numerical experimentation excludes exponential thermodynamics whereas exponential thermodynamics produces qualitatively reasonable looking genetic codes for $n_0 > 0$ whereas for $n_0 < 0$ most of the spectral power is concentrated at $p = 2$. For small values of n_0 and r_0 purely bosonic thermodynamics fails to reproduce codes satisfying the necessary conditions $D(p) > 0$ and $D(p) < 7$ satisfied by the real genetic code. Super symmetric variant with $S_p(n) = S_{B,p}(n) + S_{F,p}(n)$ however yields several codes satisfying this condition when Hamiltonian is taken to be $exp(H/T) = (r + r_0)^{n_0}$, r the number of summands in the partition.

4.2.1 Basic conditions

The basic conditions on the degeneracies are following:

1. 3 values of n should correspond to stopping codons due to their non-positive or negative entropy. Non-positive entropy is certainly the logical option since the notion of zero information codon does not seem to be reasonable. Numerical computations demonstrate that negative entropies are rather rare whereas $S_p(n) = 0$ occurs rather often. The reason is that if partition function is not divisible by any $p \leq 61$ then the smallest prime $p \leq 61$ not dividing any of the numerators of Boltzmann weights minimizes information and gives $S_p(n) = 0$. These observations suggest that $S_p(n) \leq 0$ condition should be used as a criterion for stopping codon property.

2. The degeneracies $D(p)$ satisfy $D(p) > 1$ if one has $(0,1) \to (0,1)$ so that 0 and 1 correspond to the two amino-acids coded by single DNA.

3. Complete hit means that the numbers $N(k)$ of DNAs coding $D = k \in \{2,3,4,5,6\}$ real amino-acids (as distinguished from stopping sign) should be $(9,1,5,0,3)$. This condition combined with the condition $N(stop) = 3$ allows an automatic search of candidates for codes.

4.2.2 Results

For polynomial thermodynamics the range $n_0 \in \{1,5\}, r_0 \in \{0,5\}$ is scanned. For exponential thermodynamics the range studied is $r_0 \in \{1,5\}, s_0 \in \{1,5\}$. B, F, and BF variants are studied applying the two alternative criteria for the stopping codon and requiring that exactly 3 stopping codons result.

1. $S \leq 0$ *as a criterion for the stopping codon*

a) Polynomial thermodynamics.

i) Numerical experimentation shows that the number of stopping codons increases rapidly with the values of (n_0, r_0) for polynomial thermodynamics so that only small parameter values seem to be worth of considering.

ii) For cases B and F no solutions are found. BF allows single solution. This code corresponds to $(n_0, r_0) = (2,4)$ having degeneracies

$$(D(2), D(3), ..., D(61)) = (4,4,9,3,7,6,2,2,1,1,4,1,3,1,2,2,3,4) \ .$$

The numbers of DNAs associated with the degeneracies (1, 2, 3, 4, 5, 6) are

$$(N(1), N(2), N(3), N(4), N(5), N(6)) = (6,4,3,3,0,1)$$

to be compared with the degeneracies

$$(2,9,1,5,0,3)$$

of the real code. If 3 DNAs from 9-plet ($p = 5$) and 1 DNA from 7-plet ($p = 11$) are shifted to 4 1-plets, and 1 DNA from 3-plet is shifted to 3-plet, correct degeneracies result. A modification of r_0 by adding the product of primes $p \leq 61$ with $p \notin \{5, 11\}$ would affect the degeneracies associated with 5 and 11.

b) Exponential thermodynamics.

There are no solutions for F and BF. B gives solution $(r_0, s_0) = (5,3)$ with degeneracies

$$(9,1,3,1,5,2,3,2,4,4,4,2,3,2,2,3,1,8) \ .$$

From this solution it is possible to construct the real genetic code by shifting 3 codons from 9-plet to 3 1-plets, one codon from 3-plet to a second 3-plet, and 2 codons from 8-plet to 5-plet and 3-plet.

2. $S < 0$ *as a criterion for the stopping codon*

1. Polynomial thermodynamics.
   For F and BF no solutions are found. B gives single solution $(n_0, r_0) = (3, 1)$. The degeneracies are $(3, 2, 11, 6, 3, 1, 5, 1, 5, 6, 4, 1, 2, 1, 1, 2, 2, 3)$ and quite far from those of the real genetic code.

2. Exponential thermodynamics.
   No solutions are found.

The conclusion is that BF for $S_p < 0$ criterion for stopping codon is the most realistic one and might produce by a small deformation the real genetic code.

4.3 Could Small Perturbations Of Hamiltonian Cure The Situation?

The troubling outcome of calculations is that no realistic code is found for the simplest Hamiltonian. The obvious guess is that one should study small perturbations of the Hamiltonian. There are two kinds of small perturbations. The perturbations of the first kind are small in the real sense but can induce dramatic changes of the genetic code by affecting the p-adic norms of $Z(n)$. The perturbations of the second kind are small in the number theoretical sense but as a rule affect strongly the values of the real probabilities.

4.3.1 Small perturbations in the real sense

The perturbations which are small in the real sense would simply modify $f(r)$ by few units. They would however dramatically affect the p-adic norms of $Z(n)$ and induce thorough changes in the genetic code. In order to proceed in a rational manner some additional assumptions are necessary and therefore this approach will be left in the next subsection where the maximization of the total negentropy of the genetic code is introduced as a variational principle allowing to fix the Hamiltonian as a small perturbation reducing the values of $f(r) = r$ of the Boltzmann weight. It seems that this approach is the more promising one.

4.3.2 Number theoretically small perturbations

The small values of n_0 and r_0 plus unsuccessful searches for $n_0 > 5$ encourage to ask whether the real code result from the semi-realistic codes via a small perturbation of Hamiltonian changing only the partition function Z in number theoretical sense.

The simplest situation is achieved if perturbations do not distinguish between partitions with the same value of r. The number theoretical generalization for the notion of symmetry of action principle suggests that perturbations should leave invariant the prime power factors p^k of Z for $p \leq 61$ but affect them for $p > 61$. This would affect only the probabilities of individual partitions and the positive contributions to $S_p(n)$ coming from the numerators of Boltzmann weights. This might be enough to affect the situation in the case that two primes p_1 and p_2 have nearly the same value of $S_p(n)$ in the original situation. What would be needed that three singly (and thus rarely) coded primes would become doubly coded by this kind of fine tuning.

More precisely, the Boltzmann weight associated with r transforms in $H(r) \to H(r) + \Delta H(r)$ as $B(r) = exp(-H(r)/T) \to B(r) \times (1 + \Delta H(r)/T)$. From this it is clear that the p-adic norm of the contribution of r to the partition function is unaffected if $H(r)$ is divisible by a sufficiently high powers of all primes $2 \leq p \leq 61$: this by the way defines what the notion of small perturbation means number theoretically. Obviously this kind of symmetries exist and since large powers of p in $\Delta H(r)$ modify dramatically the probabilities $p(n, r)$, it is indeed possible to affect the degeneracies associated with various amino-acids.

The simplest perturbation corresponds to the addition of a sufficiently high power of the product $P = \prod_{p \leq 61} p$ to r_0: $r_0 \to r_0 + P^k$. The p-adic norms appearing as arguments of logarithms would remain invariant. Boltzmann weights would be identical in an excellent approximation as for infinite temperature limit so that the probabilities $p(n, r)$ would reduce to $p(n, r) \simeq d(n, r)/d(n)$. The model would result via the replacement of $p(n, r) \to d(n, r)/d(n)$ from the original model.

It turns out that this replacement does not solve the problem in the range $(n_0 \in \{1, 5\}, r_0 \in \{0, 5\}$: no codes with 3 stopping sign codons are found. One cannot of course exclude the possibility that larger values of n_0 and r_0 might provide a solution.

A more general trial would assume that the perturbation modifies the p-adic norms of Boltzmann weights but leaves the norms of partition function invariant.

4.3.3 Should one break the symmetry between partitions with same r?

A more radical modification results if the perturbation distinguishes between partitions with different values of r. It is however not clear whether integer valued perturbation can be small in number theoretic sense. Rank and crank distinguish between partitions with same r.

The most radical option is to replace r with a new invariant. If rank and crank define the entire Hamiltonian, they divide partitions into equivalence classes by combining partitions with different values of r to single equivalence class so that the situation changes dramatically. The knowledge about the numbers of partitions in corresponding equivalence classes plus values of these invariants would make it easy to check whether either of them could reproduce the real genetic code.

4.4 Could One Fix Hamiltonian $H(R)$ From Negentropy Maximization?

Numerical calculations suggests that number theoretically small modifications might not be the correct manner to find a correct genetic code: the codes having the correct number of stopping codons and coding for all primes differ simply too much from the real code. Even if such a code could be found one can argue that it is only a skillful exercise in the modular arithmetics. Numerical difficulties are also obvious since powers of the product $P = 2 \times 3.... \times 61$ must be added to $f(r)$.

For the perturbations of $f(r) = r$ which are small in the real sense numerical control is not lost but with physicist's intuition in the number theory the modifications of the genetic code are completely unpredictable. The reduction of $f(r)$ by a single unit for single sufficiently small value of r could change the whole biology! In order to study them one should have additional principle allowing to get grasp of the problem.

The great principles of physics are variational principles and Negentropy Maximization Principle is the basic principle in TGD inspired theory of consciousness [K6]. Quantum criticality predicts a Universe able to engineer itself and this suggests that the Hamiltonian $H(r)$ determining the genetic code could be a result of "genetic engineering" maximizing the total negentropy of the genetic code.

4.4.1 Could one engineer $H(r)$ from the real genetic code in the case of polynomial thermodynamics?

The most general hypothesis would be that the 62 values of $f(r) = exp(-H(r)/T)$ are completely free positive integers and look whether it is possible to find a Hamiltonian reproducing the genetic code. The naïve idea is that since the number of integers $f(r)$ is the same as the values of n, a judicious choice of $f(r)$ could allow to assign to a given n arbitrary $p(n)$ or make it a stopping sign codon. If each r is shifted by the same sufficiently large power of $P = \prod_{p \leq 61} p$, the probabilities for partitions are in an excellent approximation identical in the case of polynomial thermodynamics so that the situation would reduce to a mere modular arithmetics.

One could start from $n = 2$ and proceed by increasing n and determining the value of $f(r = n)$ at n: th step from the requirement that the desired value of p results. What seems obvious is that the value of the partition function $Z(n) = \sum_{r=1}^{n} d(n,r) f(r)$ can be fixed to have an arbitrary prescribed value and thus also the $k_p(Z(n))$ giving the negative contribution to the entropy can be fixed to a desired value. This leaves still some freedom to arrange the value of $k_p(d(n,n)f(n)) = k_p(f(n))$ making possible fine tuning in the n: th numerator contributing to the entropy. The entropies $S_p(n+1)$ and $S_p(n)$ would be related by the condition

$$\begin{aligned} \frac{S_p(n+1) - S_p(n)}{log(p)} &= -k_p(Z(n+1)) + k_p(Z(n)) \\ &+ \sum_{r=1}^{n} [d(n+1,r) - d(n,r)]\, k_p(f(r)) + k_p(f(n+1)) \ . \end{aligned} \tag{4.7}$$

The modular arithmetics is of course different from real analysis and the situation might not be so simple as it looks. On the other hand, if this picture is correct, one might interpret the freedom to construct the genetic code almost at will as the fruit of quantum criticality making possible genetic engineering.

4.4.2 Maximization of the total negentropy of the genetic code as a way to fix the Hamiltonian

The basic objection to this approach is that it is not predictable. It is however possible to introduce a natural variational principle. The maximization of the total negentropy $N_{tot} = -\sum_n S_{p(n)}(n)$ of the genetic code subject to the constraint that all primes are coded and there are 3 stopping codons would in principle allow to fix the function $f(r)$ uniquely.

The maximization of the total negentropy allows to conclude that large (small) prime powers correspond to large (small) DNA multiplets. For instance, if only first powers of p appear, 9 doublets would correspond to $p = 2, ..., 23$, triplet to $p = 29$, five 4-plets to $p = 31, ..., 47$, and 3 6-plets to $p = 53, 59, 61$. Furthermore, since the value of $Z(n)$ increases with n, and thus also the probability that it has large prime power factors, one expects that large values of n should correspond to large values of p. Hence the orderings of multiplet sizes, primes powers p^k appearing as factors of $Z(n)$, and integers n should correlate strongly.

Is there then any bound on the exponents of powers p^k appearing in $Z(n) = \sum_r d(n,r)f(r)$? If not, then the variational principle does not work. For instance, one might think that ones has

$$f(n+1) = \sum_{r=1}^{n} d(n+1,r) + mp^k \quad ,$$

where k is an arbitrarily large power of p so that $Z(n+1) = mp^k$ holds true and gives an unbounded contribution to $k_p(Z(n+1))$. Only $p(n+1, n+1)$ would differ significantly from zero and would be near 1 but this does not give any restriction. It would seem that there must exist some natural bound on the values of $f(r)$ to stabilize the variational principle.

The most natural option is modulo n+1 arithmetics based on the assumption that Boltzmann factors depend on both n and r and one has $f(n,r) \leq n$ at level n. Boltzmann factors would formally restrict the partition of any integer $m > n$ to partitions of n. This would make the problem numerically more tractable. With this assumption the model for $f(r) = r$ would correspond to the maximum value of $Z(n)$. There would be $61! \sim 5 \times 10^(83)$ alternatives to be scanned but reasonable assumptions should reduce considerably this number.

One can imagine two kinds of additional assumptions.

1. If the genetic code has resulted as a product of singlet and doublet codes then one could argue that also $n = 4$ and $n = 16$ should maximize their total negentropy and code for all primes $p < n$ as real or stopping codons.

2. A much stronger additional assumption that a genetic code coding all primes $p \leq n$ results for every value of n does not work since it implies that highest primes are coded only once.

Consider the situation for the smallest values of n in the bosonic case. For $n = 2$ $f(r) = r$ implies $Z(2) = 3$ giving $p(2) = 3$ favored by local negentropy maximization and $S(2) = -log(3)$. $f(2) = 1$ would give $p(2) = 2$ and $S(2) = -log(2)$. For $n = 3$ to $f(r) = r$ would give $Z(3) = 6$ giving $p(2) = p(3) = 3$ and $S(3) = -log(3) + log(2)/3$ and $S_{tot} = -2log(3) + log(2)/3$. This corresponds to the maximum of total negentropy for 4-code. The code is consistent with the proposal that $2n$ and $2n+1$ code for the same amino-acid for $n < 61$ explaining the fact that almost all amino-acids are coded by an even number of codons. The absence of stop codon does not allow this code as a genuine singlet code. For $(f(1), f(2), f(3)) = (1, 2, 2)$ with $(Z(2), Z(3)) = (3, 5)$ one would have $n(2) = 3$ and $n = 3$ would represent stopping codon.

For larger values of n a convenient starting point would be $f(n) = n$ and direct checking of values $f(n) = n - k$ for not too large values of k to find a value of Z corresponding to a large prime power. This would give a precise content to what a small perturbation of the Hamiltonian $H(r) = log(r)$ in real sense means in practice. Perturbation would be small only in sense of real analysis and number theoretic effects would be rather dramatic for perturbations at small values

of r. Also the notion of a small perturbation of a given genetic code makes also sense. If $f(r)$ is changed only for the values of r near to $r = 63$, only the degeneracies of the amino-acids coded by largest integers and thus having largest degeneracies are affected.

4.4.3 Bosonic Hamiltonian maximizing negentropy subject to constraints coming from the real genetic code

The direct computational search of genetic codes maximizing the total negentropy without any assumptions about genetic code besides non-degeneracy requires a considerable computational power. It is much easier to search for $n \to p(n)$ assignments maximizing the negentropy subject to the constraint that the assignment is consistent with the genetic code.

The reason is that one can imagine a very simple method giving hopes of finding an assignment $n \to f(n)$, $1 \leq f(n) \leq n$ consistent with the genetic code. The basic observation is the variation of $f(n)$ in the allowed range allows always to achieve the condition $Z(n)\ mod\ p = 0$ for $p \leq n$. This gives reasonable hopes that the nearest prime $p \leq n$ maximizes $Z(n)$. Of course, it can happen that some prime $p > n$ divides $Z(n)$ or some large power of small prime divides $Z(n)$. The optimistic guess for the assignment is simple to construct by starting from $n = 63$ and by proceeding downwards in this manner. One might argue that the ansatz is too conservative. With some good luck it might be possible to assign 6-plets to quite many large primes since the probability $P(n, p)$ to find a value of $f(n)$ guaranteeing $Z(n)\ mod\ p = 0$ for p slightly larger than n is $P(n, p) = n/p$ and near to one. The assignment of 2-plets and stopping to small primes also helps to maximize the total negentropy.

Computational testing of various ansätze based on guesses for stopping codons required to correspond to as small integers as possible is rather straightforward when one starts from a simple guess deduced by the strategy above and described in table 4 below. The strategy is following.

1. It is easy to deduce the map $n \to p(n)$ for $n \leq 13$. For larger values of n prime divisors larger than that implied by the ansatz produce trouble so that the natural strategy is to look whether stopping codons could correspond to integers above $n = 14$ and near to it.

2. The larger the values n of integers corresponding to stopping codons are, the larger the numbers of values $f(n(stop))$ satisfying the criterion for the stopping codon are. The criterion is the indivisibility of $Z(n(stop))$ by any $p \leq 61$ so that prime values of $Z(n(stop))$ certainly satisfy the constraint for $n(stop) > 7$. This increases the hopes that the constraints from the real code can be satisfied. The smallest values of $n(stop)$ for which the constraints can be satisfied for all values of n are $n(stop) \in \{14, 15, 17\}$. The computation proceeds simply by checking whether any combination of candidates for these three stopping codons satisfies is consistent with the genetic code.

In the sequel considerations are restricted to the bosonic partition function but the generalization to the supersymmetric case is straightforward. **Table 8** represents the ansatz which served as a starting point for the calculations.

Maps $n \to p(n)$ consistent with the real code can be found by a numerical experimentation starting from the simple guess summarized by Table 4 above and changing the assignments in a obvious manner in the case that some relatively small n yields a large prime factor. In this manner for instance $p = 61$ multiplet can be completed to a 6-plet. The table below represents such a map. From the table it is clear that $p(n) = p(n + 1)$ symmetry is only slightly broken and can be understood as a direct consequence of the mechanism assigning to given n the desired prime $p \sim n$.

The Boltzmann weights for the $n \to p(n)$ correspondence represented in the table are given in the array below.

n in range	is coded to	multiplet
63-61	61	3
60-59	59	2_1
58-53	53	6_1
52-47	47	6_2
46-43	43	4_1
42-41	41	2_2
40-37	37	4_2
36-31	31	6_3
30-29	29	2_3
28-25	23	4_3
24-21	19	4_4
$\{20-18, 16\}$	17	4_5
17		*stop*
15-14		*stop*
13-12	12	2_4
11-10	11	2_5
9-8	7	2_6
7-6	2	2_7
5-4	5	2_8
3-2	3	2_9

Table 8: The $n \to p(n)$ correspondence whose deformation produces an $n \to f(n)$ correspondence consistent with the real genetic code.

n in set	is coded to p	multiplet
$\{63-60, 52, 27\}$	61	6_1
$\{59, 57\}$	59	2_1
$\{58, 56-53, 51\}$	53	6_2
$\{50, 49, 31\}$	47	3
$\{48-44, 33\}$	43	6_3
$\{43, 28, 26, 24\}$	23	4_1
42-41	41	2_2
40-37	37	4_2
36-32	31	4_3
30-29	29	2_3
25-21	19	4_4
$\{20, 19, 18, 16\}$	17	4_5
$\{17, 15, 14\}$		$stop_3$
13-12	13	2_4
11-10	11	2_5
9-8	7	2_6
7-6	2	2_7
5-4	5	2_8
3-2	3	2_9

Table 9: The $n \to p(n)$ correspondence maximizing the total negentropy with a constraint for multiplicities coming from the real genetic code.

n	1	2	3	4	5	6	7	8	9	10	11	12
$f(n)$	1	2	3	2	2	2	1	4	6	5	1	12
n	13	14	15	16	17	18	19	20	21	22	23	24
$f(n)$	7	12	12	3	16	4	13	19	9	18	21	18
n	25	26	27	28	29	30	31	32	33	34	35	36
$f(n)$	22	11	6	1	6	23	19	17	15	22	34	5
n	37	38	39	40	41	42	43	44	45	46	47	48
$f(n)$	11	32	32	25	41	28	10	37	35	25	11	39
n	49	50	51	52	53	54	55	56	57	58	59	60
$f(n)$	1	11	24	22	2	5	47	39	9	25	1	48
n	61	62	63									
$f(n)$	21	15	20									

4.5 Could The Symmetries Of The Genetic Code Constrain Number TheoreticalThermodynamics?

The number theoretic approach alone leaves completely open the correspondence between DNA triplets and integers n and only the comparison of a code predicting correctly the degeneracies of various amino-acids with the real genetic code allows to deduce information about this correspondence. For instance, 0, 1 DNAs and amino-acids can be identified immediately.

The model for prebiotic evolution [K5, **?**] relies on the fact that the genetic code has an almost exact symmetry: the third nucleotide of the codon has symmetry under A-G exchange and slightly broken symmetry under T-C exchange and an interesting possibility is that this symmetry could be understood at the number theoretical level. Certainly it cannot be a property of the map mapping DNA triplets to integers alone.

4.5.1 What exact A-G symmetry and almost exact T-C symmetry could mean number theoretically?

The most natural interpretation for A-G and T-C symmetries of last nucleotide of codon is that the third 4-digit of the DNA triplet interpreted as a number in the set $\{0, 63\}$ represented in 4-base do not matter much. The symmetry for T-C is slightly broken and this gives 64-20 code instead of $64 \to N \leq 16$ code. Real mathematics would suggest that these 4-digit corresponds to zeroth power of 4 whereas 2-adic arithmetics suggests that it corresponds to the second power of 4.

The characteristic feature of the genetic code is that the degeneracies come in pairs which can be understood in terms of A-G symmetry. There are 3 6-plets, 5 4-plets, 9 2-plets and 1 3-plet of amino-acids and one 3-plet of stopping codons besides the 2 singlets assignable to 0 and 1. That almost all multiplets contain even number of DNAs reflects the additional approximate T-C symmetry.

Even degeneracies must correspond to an approximate symmetry of the partition function. This kind of symmetry could be produced by hand by expressing the partition function as a product of partition functions $Z(n)$ and $Z(f(n))$, where $n \to f(n)$ represents the symmetry but numerical experimentation shows that this does not work. The reason is that for a given n the primes associated with n and $f(n)$ compete and product partition function selects winners from these pairs reducing the degeneracies of the losers so that spectral power tends to get peaked. Hence the product model works only if the symmetry is already there in the sense that the largest prime power factor for $Z(n)$ and $Z(f(n))$ correspond to same prime p.

Suppose that 3 codons correspond to stopping codons. Suppose that there exist a symmetry $n \to f(n)$, not necessary reflection, acting on remaining codons with the property that the largest prime power dividing $Z(n)$ and $Z(f(n)$ corresponds to the same p. Since the number of these codons is odd, the map $n \to f(n)$ must have a fixed point. Obviously the degeneracies are even for primes coded by non-fixed points and odd for those coded by fixed points and the structure of genetic code is consistent with this prediction.

Quite generally, for the reduction of the code to a maximal subset of integers $2 \leq n \leq 63$ not containing $f(n)$ for any n, one would have 11 singlets, 5 2-plets, and 3 3-plets in the set of even integers, or briefly

	A	G	T	C	
A	phe	ser	tyr	cys	A
	phe	ser	tyr	cys	G
	leu	ser	stop	stop	T
	leu	ser	stop	trp	C
G	leu	pro	his	arg	A
	leu	pro	his	arg	G
	leu	pro	gln	arg	T
	leu	pro	gln	arg	C
T	ile	thr	asn	ser	A
	ile	thr	asn	ser	G
	ile	thr	lys	arg	T
	met	thr	lys	arg	C
C	val	ala	asp	gly	A
	val	ala	asp	gly	G
	val	ala	glu	gly	T
	val	ala	glu	gly	C

Table 10: Genetic code.

$$29 = 10 \times \mathbf{1} \oplus 5 \times \mathbf{2} \oplus 3 \times \mathbf{3} \ .$$

The fixed point $n = f(n)$ would code the amino-acid (ile) coded by 3 DNAs.

A reasonable candidate for the symmetry is suggested by the preceding construction reproducing the degeneracies of the genetic code correctly and predicting that n and $n+1$ tend to code the same amino-acid.

A further input is the information provided by the deviations from the universality of the genetic coded to be discussed later. The deviations from the universality typically involve stopping codons and in the proposed construction it is easy to perform small modifications of the code for values of n near 63. Hence it is natural to test the stronger assumption $2n$ and $2n+1$ code for the same p for $2 \leq n < 60$ and that $n = 61, 62, 63$ act as stopping codons so that $n = 60$ would correspond to the fixed point coding for ile.

An objection against this hypothesis is that a lot of negentropy is lost if large integers are forced to act as stopping codons. Also the successful construction of the genetic code table starting from the A-G and T-C symmetries of the code table leads to the assignment of the stopping codons to relatively small integers. In this construction the assignment of stopping codons to large values of n codons would also make more difficult to assign large multiplets to large primes.

4.5.2 How close is the correlation between the map from DNA triplets to integers and the map $n \to p(n)$?

The number theoretical model alone does not fix the map between DNA triplets and integers although it poses constraints on this correspondence. A-G symmetry and almost T-C symmetry of the code table however suggest a labelling of the codons which in good approximation could determine *codon* $\to n$ map.

1. A-G and T-C symmetries suggests that the numbering of genetic codons using 4-base representations, that is as sequences of integer triplets (n_1, n_2, n_3), $0 \leq n_i \leq 3$ in 4-base such that each integer labels the four bases. The correspondence can be different for the different members of the triplet. The natural correspondence would be such that (n_1, n_2, n_3) interpreted as 4-digit representation of n gives the map in a reasonable approximation.

2. The correspondence $(T, C, A, G) \leftrightarrow (0, 1, 2, 3)$ for the third nucleotide turns out to be most realistic one from the point of view of $n \to p(n)$ correspondence. A-G and T-C symmetries

	A	G	T	C	
A	phe (46, 43)	ser(62, 61)	tyr(16, 17)	cys(31, 29)	A
	phe(47, 43)	ser(63, 61)	tyr (17, 17)	cys(32, 29)	G
	leu (44, 41)	ser (61, 61)	stop (14)	stop(30)	T
	leu (45, 41)	ser(62, 61)	stop(15)	trp(0/1)	C
G	leu(42, 41)	pro(58, 59)	his(12, 13)	arg(28, 23)	A
	leu (43, 41)	pro(59, 59)	his(13, 13)	arg(29, 23)	G
	leu(40, 41)	pro(56, 59)	gln(10, 11)	arg(26, 23)	T
	leu (41, 41)	pro(57, 59)	gln(11, 11)	arg(27, 23)	C
T	ile (38, 37)	thr(54, 53)	asn(8, 7))	ser(24, 61)	A
	ile(39, 37)	thr(55, 53)	asn (9, 7)	ser(25, 61)	G
	ile (37, 37)	thr(52, 53)	lys (6, 2)	arg(22, 23)	T
	met (1/0)	thr(53, 53)	lys (7, 2)	arg(23, 23)	C
C	val(35, 31)	ala(50, 47)	asp(4, 5)	gly(20, 19)	A
	val(36, 31)	ala(51, 47)	asp(5, 5)	gly(21, 19)	G
	val(33, 31)	ala(48, 47)	glu(2, 3)	gly(18, 19)	T
	val (34, 31)	ala(49, 47)	glu(3, 3)	gly(19, 19)	C

Table 11: Genetic code with the proposed correspondences between DNA triplets with integers n and amino-acids with primes $p(n)$. For instance, $ala(50, 47)$ tells that CGA is mapped to $n = 50$ and ala corresponds to the prime $p = 47$.

suggests that n_3 is mapped almost as such to the third 4-digit of n apart from symmetry breaking due to the complications caused by the insertion of 0 and 1 to the code table.

3. The previous example for the genetic code suggests that $n = 14, 15$ correspond to stopping codons. Negentropy Maximization Principle favors doublets for small integers. Since the third column of the code table contains only doublets, it should correspond to small integers. These constraints are satisfied under two conditions. First, n_1 labels the rows of the table with the correspondence $(T, C, A, G) \leftrightarrow (0, 1, 2, 3)$ along the rows of the table so that first and second and third and fourth columns are permuted. Secondly, n_2 must label the entries formed by 4-sub columns of the table and one must have $(C, T, G, A) \leftrightarrow (0, 1, 2, 3)$ for so that n_2 increases from bottom to top.

4. The two stopping codons ATT and ATC would would correspond to $n = 14, 15$ as in the example discussed above. Stopping codon ACT would correspond to $n = 30$ ($n = 17$ in the example). Encouragingly, ser corresponds to $\{63, 62, 61, 60, 22, 23)\}$ coding very naturally $p = 61$. Also in the example discussed 61 belongs to 6-plet.

5. The correspondence between codons (n_1, n_2, n_3) and integers n cannot be given exactly by the representation of n in 4-base since 0 and 1 do not correspond to ACC coding trp (0 or 1) but would correspond to $(1, 3, 3) = 31$. TAC coding met (1 or 0) would correspond to $(2, 1, 3) = 39$. The map from codons to integers with minimal symmetry breaking is obtained from the 4-digit coding of n by shifting 0 and 1 to the positions of 31 and 39. For $n(codon) < 29$ this induces the map $n = n(codon) + 2$. For $41 > n(codon) > 29$ the map is $n = n(codon) + 1$, and for $n(codon) > 41$ the map is $n = n(codon)$. Table 7. lists the resulting number theoretic code in the bosonic case and its correspondence with DNA triplets and amino-acids for this option. It is clear that the risky assignments $n \to p(n)$ are associated with ser and pro.

1. Numerical testing in the bosonic case

It is straightforward to test the proposed $n(codon) \to n$ map by numerical computations. They are done for both bosonic and supersymmetric case. In the bosonic case correspondence cannot be realized as such. $n = 29$ corresponding to 6^{th} arg is the source of problems and by a trial and error one ends up with a slightly modified $p \to n(p)$ correspondence allowing two solutions for

n	1	2	3	4	5	6	7	8	9	10	11	12	13
model	0	3	3	5	5	2	2	7	7	11	11	13	13
real	0	3	3	5	5	2	2	7	7	11	11	13	13
$f_1(n)$	1	2	3	2	2	2	1	4	6	5	1	12	7
$f_2(n)$	1	2	3	2	2	2	1	4	6	5	1	12	7
n	14	15	16	17	18	19	20	21	22	23	24	25	26
model	0	0	17	17	19	19	19	19	23	23	61	61	23
real	0	0	17	17	19	19	19	19	23	23	61	61	23
$f_1(n)$	4	8	6	10	12	12	3	5	12	11	15	17	7
$f_2(n)$	4	12	2	6	12	12	7	5	16	11	15	17	7
n	27	28	29	30	31	32	33	34	35	36	37	38	39
model	23	23	0	29	29	23	31	31	31	31	37	37	43
real	23	23	0	29	29	23	31	31	31	31	37	37	37
$f_1(n)$	7	6	21	28	20	27	33	33	21	21	15	36	30
$f_2(n)$	3	6	21	24	24	31	33	33	17	17	15	32	26
n	40	41	42	43	44	45	46	47	48	49	50	51	52
model	37	41	41	41	41	41	41	43	47	47	47	47	53
real	41	41	41	41	41	41	43	43	47	47	47	47	53
$f_1(n)$	13	6	18	23	12	27	28	44	24	33	14	23	52
$f_2(n)$	9	10	18	23	12	31	32	5	20	29	6	23	10
n	53	54	55	56	57	58	59	60	61	62	63		
model	53	53	53	59	61	59	59	59	61	61	61		
real	53	53	53	61	59	59	59	59	61	61	61		
$f_1(n)$	29	26	25	56	10	49	9	10	4	27	49		
$f_2(n)$	5	48	45	41	14	56	15	27	8	35	22		

Table 12: $n \to p(n)$ correspondence allowing two different Boltzmann weights $f_1(n)$ and $f_2(n)$ consistent with the real genetic code obtained by a small modification of the correspondence $n \to p(n)$ implied by the map $n(codon) \to n$ discussed above. This correspondence is also shown in the table for comparison purposes.

Boltzmann weights $f(n)$ represented in Table 8 below. The requirement that the small deviations from the standard code are realizable as small deviations of $f(n)$ without affective genetic code leaves only $f_1(n)$ into consideration (as will be found in the next section).

2. Numerical testing in the supersymmetric case

One might hope that the replacement of the bosonic partition function with the super-symmetric one might allow an exact realization of the simplest $n(codon) \to n$ correspondence. The multiplication of Z_B by Z_F does not destroy any divisors already present so that the effect might be small. It however turns out that the proposed ansatz fails already at $n = 10$ since Z_F equals to prime $p = 23$ giving higher negentropy than $p = 11$-factor of Z_B. One can try to continue by a modification of the ansatz but the troubles continue and are basically due to the large prime power factor of Z_F. Hence it seems that bosonic ansatz is the only realistic one. Fermionic ansatz is certainly non-realistic since the number of non-vanishing elements of $d_F(n, r)$ is as small as 10 even for $n = 63$.

5 Confrontation Of The Model With Experimental Facts

The proposed model of genetic code means that we would rather literally consist of sequences of numbers with DNA representing sequences in base 64 and amino-acid sequences represented as products of primes $2 \leq p \leq 61$ and separated by zeros. What this predicts depends on how literally we take this interpretation.

5.1 Basic Facts About Amino-Acids

Amino-acids can be classified into three groups.
i) The first class contains 8 hydrophobic non-polar amino-acids with non-polar neutral side-chain. They are leu (6), ala (4), val (4), pro (4), ile (3), phe (2), met (1), trp (1) (numbers in parenthesis tell the number of DNAs coding the amino-acid in question).
ii) Second class consists of 7 hydrophilic polar amino-acids with polar neutral side-chain: ser (6), gly (4), thr (4), cys (2), asp (2), gln (2), tyr(2).
iii) The third class consists of polar hydrophilic acidic amino-acids with charged side chain: asp (2), glu (2) and hydrophilic basic amino-acids arg (6), lys (2), his (2): 5 altogether.

As already noticed, met and trp representing 0 and 1 should belong to the group of non-polar neutral amino-acids and indeed do so. Also the amino-acid representing a fixed point of symmetry $n \to f(n)$ (ile) (if such a symmetry indeed exists) would belong to this group. It is worth of noticing that each group contains single amino-acid coded by 6 DNAs.

5.2 Could The Biological Characteristics Of An Amino-Acid Sequence Be Independent On The Order Of Amino-Acids?

The representation of an integer as a product of primes does not depend on the order of factors. Unless the amino-acid sequence does not inherit the natural order of DNA triplets somehow, the biological properties of portions of amino-acid sequences separated by zeros would be invariant under the permutations of amino-acids: the permutation of amino-acids would be analogous to a permutation of bosons. The prediction is extremely strong and certainly testable and might have been observed long ago if indeed true. Professional biologist could probably immediately kill this option.

5.3 Are The Amino-Acids And DNAs Representing 0 And 1 Somehow Different?

The amino-acid representing 0 would most naturally separate different structural and/or functional units and both 0 and 1 could represent a biologically inert amino-acid. Also other interpretation might of course be imagined. The amino-acids representing 0 and 1 would be met coded by TAC and trp coded by ACC, not necessarily in this order.

Do met and trp then have some special properties distinguishing them as 1 and 0?

1. Consider first chemical structure. Both are neutral and non-polar amino-acids, which can be regarded as a basic prerequisite for biological inertness. Met is the only amino-acid containing $CH_2 - S - CH_3$ side chain (cys contains $CH_2 - S - H$ side chain and there are no other amino-acids containing sulphur). Trp in turn is the only amino-acid containing two cyclic chains.

   This kind of arguments must be however taken with a big grain of salt as the following argument shows. Proline differs from all other amino-acids in that the neutral group $H_3N^+ - COO^- - C - H$ group is replaced by a charged $H_2N - COO^- - C - H$ group and is therefore a reasonable looking candidate for 0: pro is however coded by 4 DNAs which would correspond to 2 n and $n + 2$ DNAs with $2 \leq n \leq 31$.

2. At the level of biological function there is indeed a deep difference. The DNA triplet coding for met acts almost universally (for deviations see [I1]) as a starting codon which conforms with the identification of met as an amino-acid representing either 0 or 1 (literally the first amino-acid!) and having no other biological significance than telling where in a more complex structure consisting of amino-acid sequences a structural basic element coded by single gene begins.

5.4 The Deviations From The Standard Code As Tests For The Number Theoretic Model

One can take two different attitudes concerning the deviations from the universality of the code.

Since the deviations occur in mitochondrial genomes and in nuclear genomes of some unicellular eukaryotes, one could argue that in these cases the code need not have achieved full negentropy maximization and that NMP model does not apply. Even if NMP applies, one can argue that the maps $n(codon) \to n$, $n \to f(n)$, and even $n \to p(n)$ correspondence can differ dramatically from that for the nuclear genome.

Second attitude would be that these codes correspond to different local maxima of total negentropy and that the codes correspond to small perturbations of nuclear $n(codon) \to n$, $n \to f(n)$ correspondences.

The deviations from the standard genetic code [I1] allow to test between these options, in particular the genetic variant of Negentropy Maximization Principle predicting that small perturbations of $f(n)$ inducing small perturbations of genetic code can affect only large values of n. Numerical experimentation suggest that small perturbations of $n(codon) \to n$, $n \to f(n)$ correspondences are not in question.

5.4.1 Violations of universality for nuclear genes are consistent with the number theoretical model

The violations of the universality [I1] are very rare for nuclear genes. A few unicellular eukaryotes have been found that use one or two of three stop codons to code amino-acids instead. The use of two stop codons to code amino-acids necessarily violates the universality of the third column of the code table.

These violations would be consistent with the hypothesis that the two stopping codons ATA and ATG correspond to large values of n (most naturally 62 and 63) but do not force this model. For the codes represented in **Table 13** however ATA and ATG however correspond to $n = 14, 15$ so that the modification of the code occurs at rather small values of n and the modifications of $f(n)$ at these values radiate their effect to all higher values of $f(n)$ via the coupling $Z(n) = \sum_{r=1}^{n} d(n,r)f(r)$ and this effect is large in number theoretical sense. Hence small perturbations of $n \to f(n)$ and $n(codon) \to n$ correspondences might not be enough and even $n \to p(n)$ correspondence might need a modification. A detailed numerical computation is required to check whether the model can reproduced the modified codes with some assignment $f(n)$ of Boltzmann weights.

5.4.2 The mitochondrial deviations related to codons representing 0, 1, and stopping sign

For the mitochondrial genes the situation is more complex. There are several kinds of deviations and first kind of deviations related to codons representing 0, 1, and stopping sign.

1. Deviations

Consider first the exceptions associated with stopping codons and codons representing usually 0 and 1 in the proposed model.

1. Mitochondrial codon ACT from animals and micro–organisms (but not from plants) codes trp instead of stopping sign. The problem is that trp corresponds to singly coded amino-acid and should represent either 0 or 1.

2. Most animal mitochondria use TAT in the first column of the code table to code met instead of ile coded usually 3 times. Also this is troublesome since met should correspond to $n = 0$ and be coded only once.

Since both trp and met correspond to 0 and 1 in either order in the model, the question what it means that DNA not representing 0 or 1 codes for 0 or 1. The working hypothesis is that met codes for $p = 1$ whereas trp codes for 0 acting as a codon separating two functional units of amino-acid sequence and being in this sense almost equivalent with stopping codon.

1. Does the notion of $p = 1$ codon make sense?

The condition $S_{p(n)}(n) = 0$ is the most general manner to define effective $p = 1$ codon whereas stopping codon would has positive entropy. This requires that for effective $p = 1$ codons $Z(n)$ is divisible by $p(n)$ and gives a negative contribution to $S_{p(n)}(n)$ but despite this $S_{p(n)}$ is vanishing or positive.

n	27	28	29	30	31	32	33	34	35	36	37	38	39
p(n)	23	23	0	29	29	23	31	31	31	31	13	37	43
$f_{11}(n)$	7	6	21	28	20	27	33	33	21	21	21	30	24
$f_{12}(n)$	7	6	21	28	20	27	33	33	21	21	22	29	23
$f_{13}(n)$	7	6	21	28	20	27	33	33	21	21	30	21	15
$f_{21}(n)$	3	6	21	24	24	31	33	33	17	17	9	38	32
$f_{22}(n)$	3	6	21	24	24	31	33	33	17	17	10	37	31
$f_{23}(n)$	3	6	21	24	24	31	33	33	17	17	21	26	20
$f_{24}(n)$	3	6	21	24	24	31	33	33	17	17	22	25	19
n	40	41	42	43	44	45	46	47	48	49	50	51	52
p(n)	37	41	41	41	41	41	41	43	47	47	47	47	53
$f_{11}(n)$	13	6	24	23	18	27	28	44	24	27	14	23	46
$f_{12}(n)$	13	6	25	23	19	27	28	44	24	26	14	23	45
$f_{13}(n)$	13	6	33	23	27	27	28	44	24	18	14	23	37
$f_{21}(n)$	9	10	12	23	6	31	32	5	20	35	6	23	16
$f_{22}(n)$	9	10	13	23	7	31	32	5	20	34	6	23	15
$f_{23}(n)$	9	10	24	23	18	31	32	5	20	23	6	23	4
$f_{24}(n)$	9	10	25	23	19	31	32	5	20	22	6	23	3
n	53	54	55	56	57	58	59	60	61	62	63		
$p(n)$	53	53	53	59	61	59	59	59	61	61	61		
$f_{11}(n)$	29	26	25	56	10	49	15	10	4	27	55		
$f_{12}(n)$	29	26	25	56	10	49	16	10	4	27	56		
$f_{13}(n)$	29	26	25	56	10	49	24	10	4	27	3		
$f_{21}(n)$	5	48	45	41	14	56	9	27	8	35	16		
$f_{22}(n)$	5	48	45	41	14	56	10	27	8	35	17		
$f_{23}(n)$	5	48	45	41	14	56	21	27	8	35	28		
$f_{24}(n)$	5	48	45	41	14	56	22	27	8	35	29		

Table 13: $n \to p(n)$ correspondence allowing three different Boltzmann weights $f_{1i}(n)$ and 4 different Boltzmann weights $f_{2i}(n)$ as perturbations of $f_1(n)$ and $f_2(n)$. $f_{11} = f_1$ and $f_{21} = f_2$ implies that these "perturbations" are unique and correspond to $p(37) = 13$ instead of $p(37) = 37$. The table lists only the rows for which deviation from $f_1(n)$ and $f_2(n)$ occurs.

Perhaps it is not a accident that the triply coded ile corresponds to the exceptional multiplet with odd degeneracy. As proposed, single odd degeneracy could be understood if the partition function has an approximate symmetry $n \to f(n)$ such that the DNA coding the third ile corresponds to a fixed point of this symmetry. The fixed point codon would code for 1 in the sense proposed rather than for ile.

In the proposed model $p = 37$ corresponds to ile and the ile transforming to met in yeast mitochondria is coded by $n = 37$. Numerical search demonstrates that $p(37) = 13$ instead of $p(37) = 37$ provides 3 modifications of $f_1(n)$ and 4 modifications of $f_2(n)$ for which $p = 1$ condition is satisfied. The solutions $f_{11}(n)$ and $f_{21}(n)$ are identical with $f_1(n)$ and $f_2(n)$. Obviously this solution is unique.

3. What coding of $p = 0$ could mean?

What it means that $n > 0$ codes instead of stopping sign for 0 is more difficult to interpret unless 0 indeed effectively represents an amino-acid (trp) separating functionally independent units of amino-acid sequence effectively coded by separate genes and stopping sign in this sense. One might think that code has evolved like a computer program via modularization so that in the advanced form of the code DNA sequences code only for the basic building amino-acid sequences rather than their composites separated by exotic amino-acids. Other deviations are consistent with the genetic variant of Negentropy Maximization Principle.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
nuclear	0	3	3	5	5	2	2	7	7	11	11	13	13
$f_2(n)$	1	2	3	2	2	2	1	4	6	5	1	12	7
n	14	15	16	17	18	19	20	21	22	23	24	25	26
nuclear	0	0	17	17	19	19	19	19	23	23	61	61	23
$f_2(n)$	4	12	2	6	12	12	7	5	16	11	15	17	7
n	27	28	29	30	31	32	33	34	35	36	37	38	39
nuclear	23	23	0	29	29	23	31	31	31	31	37	37	43
$f_2(n)$	3	6	21	24	24	31	33	33	17	17	15	32	26
n	40	41	42	43	44	45	46	47	48	49	50	51	52
nuclear	37	41	41	41	41	41	41	43	47	47	47	47	53
yeast	53	53	41	53	53	43	43	41	47	47	47	47	53
$f_2(n)$	9	10	18	23	12	31	32	5	20	29	6	23	10
$f(n)$	13	2	18	22	10	3	36	42	14	29	4	20	7
n	53	54	55	56	57	58	59	60	61	62	63		
nuclear	53	53	53	59	61	59	59	59	61	61	61		
yeast	41	59	41	61	41	59	59	59	61	61	61		
$f_2(n)$	5	48	45	41	14	56	15	27	8	35	22		
$f(n)$	53	53	53	59	61	59	59	59	61	61	61		

Table 14: $n \to p(n)$ correspondence allowing single distribution $f(n)$ of Boltzmann weights consistent with the genetic code of yeast mitochondria obtained by a small modification of the correspondence $n \to p(n)$ implied by the map $n(codon) \to n$ discussed above. The $n \to f(n)$ correspondence results as a modification of $n \to f_2(n)$ for nuclear genetic code so that this option is favored by universality. Only the rows of $n \to p(n)$ and $n \to f(n)$ correspondences differing from those for the nuclear code are given in the table. The correspondences for nuclear genetic code are also shown in the table for comparison purposes.

5.4.3 The anomalous behavior of yeast mitochondria

Yeast mitochondria use GAX codons in the first column to code for thr (coded by 4 codons usually) instead of leu (coded by 6 codons usually). For the $n \to p(n)$ correspondences motivated by the mapping $n(codon) \to n$, the deviation would mean that the integers $n = 40 - 43$ code for $p = 53$ (thr) besides n in the range 52-55. A rough modular arithmetics based estimate for the probability that this occurs for single codon is roughly n/p for $n < p$ so that the total probability for this to occur would be $P = 40 \times 41 \times 42 \times 43/53^4 \simeq .38$. It turns out that $n = (40, 41, 42, 43)$ fails to code for $p = 53$. Thus mitochondrial code and nuclear code for yeast should have slightly different $n(codon) \to n$ correspondence. The modification

$$\begin{aligned} p(40, 41, 42, 43, 44, 45, 46, 47) = &\quad (41, 41, 41, 41, 41, 41, 43, 43) \\ \to &\quad (53, 53, 41, 53, 53, 43, 43, 41) \end{aligned}$$

is consistent with negentropy maximization. This means that the permutations $(42 \leftrightarrow 44)$ and $(45 \leftrightarrow 47)$ distinguish the map $n(codon) \to n$ from that for the nuclear code. The modification is given in **Table 14**.

5.4.4 The deviations associated with exotic amino-acids and stopping sign codons

There are also two non-standard amino-acids: selenocysteine and pyrrolysine.

1. Selenocysteine is encoded by ACT (fourth column) coding stopping sign normally. Interestingly, ACT codes also stopping sign and the translation machinery is somehow able to discriminate when selenocysteine is coded instead of stop. This codon usage has been found in certain Archaea, eubacteria, and animals. This deviation means that the number of amino-acids is 21 or 20 depending on context.

2. In one gene found in a member of the Archaea, exotic amino-acid pyrrolysine is coded by ATC, which corresponds to the lower stopping sign in the code table. This case represents the only deviation from universality of the third column of the code table but even in this case also stopping sign is coded. How the translation machinery knows whether to code pyrrolysine or to stop translation is not yet known.

   These deviations are consistent with the number theoretical models discussed in [K5, **?**] for which number 21 indeed has a deep number theoretical meaning and assuming that stopping sign can be regarded formally as an amino-acid. In the recent model a reasonable looking interpretation of the exotic amino-acids is as variants of stopping sign in some sense. For instance, the resulting amino-acid sequences could consist of separate functional units separated by selenocysteine/pyrrolysine.

To sum, all deviations challenging the number theoretic model discussed in this chapter are associated with mitochondrial genome only and involve stopping sign codons, codons representing 0 and 1 and exotic amino-acids.

5.5 Model For The Evolution Of The Genetic Code And The Deduction Of $N \to P(N)$ Map From The Structure Of tRNA

In [**?**] a detailed model for the evolution of the genetic code is developed. The hypothesis is that recent DNA-amino-acid code evolved from a code mapping RNA triplets to RNA triplets with the mediation of pre-RNA catching RNA molecules from environment and bringing them to the growing RNA sequence. Amino-acids served originally as catalyzers of the reaction but at some stage began attach to the growing RNA sequence after which RNA sequence become un-necessary and only amino-acid sequence remained.

In the recent framework tRNA would represent the mapping of integers represented by RNA as sequences in 64 base to RNAs representing sequences of primes. Genetic coded literally mapped RNA representing integer $0 \leq n \leq 63$ to an RNA representing the prime $p(n)$. The map $n \to p(n)$ could be determined up to a permutation of the 18 primes $2 \leq p \leq 61$ and permutations of integers mapped to same p (not larger than 6) from the structure of the recent tRNA since tRNA molecules could still contain RNA pairs representing $n(p) - p$ pairs. That mRNA-RNA correspondence at the level of tRNA would represent $n \to p(n)$ correspondence means that there is no need to ponder the problem how to assign to a given amino-acid the corresponding prime p: tRNA-amino-acid correspondence would be determined by biochemistry.

5.6 Genetic Code As A Product Of Singlet And Doublet Codes?

The model of the genetic code applies to any number n of DNAs and maps the numbers $n = 0, 1...n-1$ to $\{0,1\} \cup \{primes\ p \leq n-1\}$. In [**?**] a model for the genetic code resulting via a symmetry breaking from the product of codes associated with 16 DNA doublets and 4 DNA singlets was considered. At the level of DNAs the product code is very natural and the almost symmetries of the genetic code with respect to the last codon support the idea.

The product structure at the level of amino-acids is however not at all manifest and seems to be absent. This is what the number theoretical model predicts. The primes associated with the product of singlet and doublet codes have no natural composition into products of primes associated with singlet and doublet codes. Nor is the number of these primes product of numbers of primes associated with singlet and doublet codes.

6 Exponential Thermodynamics Does Not Work

In the following various unsuccessful attempts to understand genetic code in terms of exponential thermodynamics using Hamiltonian $H(r) = r$ are summarized.

6.1 What Can One Conclude About P-Adic Temperature Associated With The Genetic Code In The Case Of Exponential Thermodynamics?

Ordinary thermodynamics suggests that also in the case of exponential thermodynamics temperature should be non-negative. This would boil down to basic requirement $q_0 = r_0/s_0 > 1$ characterizing the genetic temperature. This condition has been however dropped in computations since it is not mathematically necessary in the case of finite state system.

The work with exponential thermodynamics is restricted to the bosonic case. As already found, the fermionic high temperature limit is extremely unrealistic. One important requirement is that also the primes 37 and 61 can appear as divisors in the generalization of $d(n)$ to be discussed. For the remaining primes the most conservative, and probably unrealistic, assumption would be that the arguments of the logarithms appearing in S_p are unaffected so that only the reduction of large r contributions would reduce the degeneracies of over-represented primes. It seems that for small over-represented primes the norms of logarithms must be affected.

The requirement that all entropies $S_{p(n)}(n)$ associated are negative poses strong conditions on q_0, and this might not be possible for all n. The entropic or zero entropy integers could correspond to stopping sign codons.

1. Conditions on q_0

Writing $q_0 = r_0/s_0 > 1$ one can express S_p and assuming $H = r - 1$ and $T_r = 1$ in terms of integers alone:

$$\begin{aligned} S_p(n) &= \sum_{r=1}^{n} p(n,r)(\frac{r_0}{s_0})^{-r+1} log(|\frac{r_0^{n-r+1}s_0^{r-1}}{\hat{d}(n)}|_p) \ , \\ \hat{d}(n) &= \sum_{r=1}^{n} \hat{d}(n,r) \ , \\ \hat{d}(n,r) &= r_0^{n-r+1}s_0^{r-1}d(n,r) \ . \end{aligned} \tag{6.1}$$

The use of different representation for $p(n,r)$ and the argument of logarithm is especially convenient in the numerical calculation of entropy since modular arithmetics can be applied to deduce the argument of logarithm.

To make the representation more fluent, introduce the set $\mathcal{Q}_R$ as subset of primes $p \in \mathcal{P} = \{2, 3, ..., 61\}$ by excluding primes in the set $\mathcal{R} \subset \mathcal{P}$. It turns out that $\mathcal{Q} = P \setminus \mathcal{R}$ condition is too restrictive and hence the subscript $_R$ is added to the definition. The minimal choice for $\mathcal{R}$ is $\mathcal{R}_{min} = \{37, 61\}$ but also 23 is a reasoanable candidate for an element of $\mathcal{R}$. More explicitly,

$$\mathcal{Q}_{max} = \mathcal{P} \setminus \mathcal{R}_{min} = \{2, 3, 5, 7, 11, 23, 17, 19, 23, 29, 31, 41, 43, 47, 53, 59\} \ .$$

Define also integer X_Q as the product of primes in $\mathcal{Q}$:

$$X = \prod_{p_k \in \mathcal{Q}} p_k \ . \tag{6.2}$$

Consider now the conditions on q_0 in more detail.

1. Every prime $2 \leq p \leq 61$ must divide $\hat{d}(n)$ for some values of $n(p)$ in order that the prime in question has integers n mapped to it. This has two implications. First, the arguments of the logarithms appearing in the entropy should remain invariant for all primes in $\mathcal{Q}$ to guarantee that no prime is lost. Secondly, for each prime $q \in \{23, 31, 61\}$ there should exist n_q such that $\hat{d}(n)$ is divided by q and q corresponds to the largest prime power of prime in $\hat{d}(n)$.

2. Stopping sign codons correspond to zero information integers n not containing $p \leq 61$ in their decomposition to primes. Assume that $n = 13$ and 36 remain such primes so that $\hat{d}(13)$ and $\hat{d}(36)$ remain indivisible by $p \leq 61$. Also a third similar integer must emerge in finite temperature thermodynamics.

2. Conditions for primes in $\mathcal{Q}$

Consider now these conditions for primes in $\mathcal{Q}$.

1. The p-adic norms of $\hat{d}(n,r)$ and $\hat{d}(n)$ are same as those of $d(n,r)$ and $d(n)$ if the conditions

$$r_0 \ mod \ p = 1 \ , \quad s_0 \ mod \ p = 1 \tag{6.3}$$

   hold true. This guarantees that logarithms appearing in S_p are unaffected.

2. These conditions could hold for all primes in $\mathcal{Q}$ and can be satisfied by the ansatz:

$$\begin{aligned} r_0 &= 1 + R_0 X \ , \ s_0 = 1 + S_0 X \ , \\ X &= \prod_{p_k \in \mathcal{Q}} p_k \ . \end{aligned} \tag{6.4}$$

   Note that one must have $R_0/S_0 > 1$ in order to have a positive temperature T.

3. The condition $\mathcal{Q}_R = \mathcal{P} \setminus \mathcal{R}$ is un-necessarily restrictive. One can also consider the situation in which one drops some over-represented small primes from X. The dropping of say $p = 7$ and $p = 11$ could make possible the representability of 23 appearing as a factor in $d(32) = 3 \times 11^2 \times 23$ and $d(33) = 3^2 \times 7^2 \times 23$. In fact, the dropping of all small primes $p \leq 11$ might cure at single stroke the over-representability problem. They are probably not lost totally since they have a considerable probability to appear as factors in $\hat{d}(n)$.

3. Conditions for primes in $\mathcal{R}$

Consider next the situation for a prime $q \in \mathcal{R}$, say $\mathcal{R}_{min} = \{37, 61\}$. The task is to deduce conditions on the integers (R_0, S_0).

1. There must exist at least one n_q such that $\hat{d}(n_q)$ is divisible by q:

$$\begin{aligned} \hat{d}(n_q) &= \ mod \ q = 0 \ , \ q \in \mathcal{R} \ , \\ \hat{d}(n_q) &= \sum_{r=1}^{n_q} \hat{d}(n_q, r) \ , \\ \hat{d}(n_q, r) &= (1 + R_0 X)^{n_q - r + 1} (1 + S_0 X)^{r-1} d(n_q, r) \ , \\ X &= \prod_{p_k \in S} p_k \ . \end{aligned} \tag{6.5}$$

   S_0 and R_0 satisfying these conditions for some n_q can be found by a direct numerical search.

2. For each n_q there must exist at least one r_q satisfying the condition

$$\hat{d}_{n_q r_q} \ mod \ q \neq 0 \ , \ q \in \mathcal{R} \ . \tag{6.6}$$

   These conditions are very general and allow many solutions (R_0, S_0).

   i) For $\mathcal{R}_{min} = \{37, 61\}$ and $\mathcal{Q}_{max} = \mathcal{Q}_{R_{min}}$ one can use the conditions $X_Q = X_{min} \ mod \ 37 = 7$ and $X_{min} \ mod \ 61 = 1$ to reduce conditions to a numerically more tractable form

$$\sum_{r=1}^{n_{37}}(1+7R_0)^{n_{37}-r+1}(1+7S_0)^{r-1}d(n_{37},r)\ mod\ 37 = 0\ ,$$
$$\sum_{r=1}^{n_{61}}(1+R_0)^{n_{61}-r+1}(1+S_0)^{r-1}d(n_{61},r)\ mod\ 61 = 0\ . \quad (6.7)$$

ii) If one drops the over-represented small primes $p \leq 11$ from X one obtains $X_Q = X_{min}\ mod\ 37 = 27$ and $X_{min}\ mod\ 61 = 40$. In this case conditions are obtained from previous ones by the replacement $(7,1) \to (27,40)$.

iii) For $\mathcal{R} = \{23, 37, 61\}$ one would have $X_R\ mod\ 23 = 10$, $X\ mod\ 37 = 22$ and $X\ mod\ 61 = 2$ and one would have the conditions

$$\sum_{r=1}^{n_{23}}(1+10R_0)^{n_{37}-r+1}(1+10S_0)^{r-1}d(n_{37},r)\ mod\ 23 = 0\ ,$$
$$\sum_{r=1}^{n_{37}}(1+22R_0)^{n_{37}-r+1}(1+22S_0)^{r-1}d(n_{37},r)\ mod\ 37 = 0\ ,$$
$$\sum_{r=1}^{n_{61}}(1+2R_0)^{n_{61}-r+1}(1+2S_0)^{r-1}d(n_{61},r)\ mod\ 61 = 0\ . \quad (6.8)$$

6.2 Low Temperature Limit Of Exponential Thermodynamics

The case $s_0 = 1$ ($S_0 = 0$) corresponds to integer valued q_0 and to the low temperature limit of number theoretical thermodynamics characterized by R_0 alone. In this case only $r = 1$ partition contributes significantly to $S_p(n)$ and one expects that the genetic code is determined by the decomposition of the probability $p(r = 1) = r_0^n/\hat{d}(n)$ to prime factors. The positive contribution to information comes from $\hat{d}(n)$ so that in practice this is of primary interest.

The deduction of primes minimizing $S_p(n)$ can be done conveniently by separating the calculation of the exponents of the p-adic norms from the calculation of probabilities. The calculation of the probabilities from their basic formulas is convenient due to the rapid convergence of the exponents $(1 + R_0X)^{-r+1}$ $r = 1$ term indeed gives an excellent approximation to $S_p(n)$ so that the decomposition of $\hat{d}(n)$ to primes determines $p(n)$ completely unless $d(n, r)$ compensates for the exponential decrease. This might of course mean that the assumption $S_0 = 1$ is not realistic. The study of the low temperature limit in detail can however provide valuable information about a more realistic model.

The overall idea is simple.

1. The primes in $\mathcal{R}_{min} = \{37, 61\}$ must divide $\hat{d}(n)$ for some values of n and these give conditions on R_0.

2. Sum of the over-represented small primes $n \leq 11$ can be dropped from Q and thus from X_Q to see whether $\hat{d}(n)$ is not anymore divisible by these primes so often.

The computational algorithm for finding candidates for realistic genetic codes uses the fact that the number N of DNA triplets coding given amino-acid is never large than 6 for the real genetic code.

1. Form an array of plausible looking choices of X labelling the models to be studied.

2. Calculate the allowed values of R_0 for a given model X and arrange them to a vector.

3. Calculate the components of the vector $p(n)$ for allowed values of R_0 for given X one by one. Keep count of the number of occurrence $N_n(i) = N(p(i))$ of prime $p(i)$ $i = 1, , ..18$ for given (X, R_0) as n increases. If the number $max\{N_i, i = 1, .., 18\}$ exceeds 6, stop the further scanning of n values as useless and start to test the next value of R_0.

Preliminary calculations suggest that the predictions of low temperature thermodynamics do not differ in an essential manner from those of high temperature thermodynamics. The problem is still posed by the over abundance of small primes. The reason is that in the decomposition of integer small primes are most abundant whereas large primes are rare. The probability that small prime p divide random integer is $P = 1/p$. $p = 11$ seems to be the boundary between under-represented primes and over-represented primes. Typically about 40 integers code for primes $p \leq 11$.

6.3 How To Find The Critical Temperature In Exponential Thermodynamics?

The challenge is to understand whether and how $S_0 > 0$ could cure the situation and whether there exists something analogous to a critical temperature in the sense that large long range fluctuations for ordinary criticality correspond to large degeneracies for large primes. From the point of view the association of a number theoretical critical temperature to genetic code would be rather natural since in TGD framework living systems indeed are quantum critical systems. and genetic code should be something completely exceptional.

The following arguments give some glimpse about what criticality might mean.

1. For $r_0 \sim s_0$ near criticality the probabilities $p(n,r) = r_0^{n-r} s_0^r p(n,r)/\hat{d}(n)$ are of same order of magnitude so that all values of r contribute significantly to $S_p(n)$ as in the case of infinite temperature limit. Individual contributions are however relatively small for large values of $\hat{n}$.

2. In the argument of logarithm the small primes appearing as factors of $r_0^{n-r} s_0^(r-1) p(n,r)$ tend to compensate the small primes dividing $\hat{d}(n) = \sum_r r_0^{n-r} s_0^(r-1) d(n,r)$ so that only a small number of terms with negative entropy remains and the small value of $p(n,r)$ means that overall contribution is small.

The cautious conclusion is that at criticality r_0 and s_0 should be near to each other. There are however tight constraints. For instance, for $s_0 = 1$ r_0 cannot be divisible by primes $2 \leq p \leq 61$ since in this case the partition functions would not be divisible by any of these primes and corresponding amino-acids would not be coded at all. There one must have $r_0 \geq 67, s_0 \geq 67$ in order to not lose the primes from the partition function.

The preliminary computations with small values of r_0 and s_0 near shows that realistic looking degeneracies result except for $p = 2$ whose degeneracy is of order 40 typically: it seems that the spectral power is shifted from primes $p \leq 11$ to $p = 2$. The very special character of $p = 2$ suggests a possible remedy. Perhaps the integers 0, 1, 2 should be mapped to themselves by genetic code and only odd primes compete in the variational principle. This would however mean that the number of amino-acids coded by single DNA would be 3 rather than the observed 2 consistent $(0,1) \to (0,1)$ hypothesis. This option can work only if one maps some other DNAs than 0 (1) to 0 (1). This could make sense only in the case that all primes give $S_p(n) = 0$ for some n. It turns out that the dropping of $p = 2$ only shifts the spectral power to $p = 3$ for checked small values of (r_0, s_0). It seem that if the idea of criticality is not enough unless one has clear idea about what makes (r_0, s_0) critical.

The first TGD inspired model for genetic code was based on the Combinatorial Hierarchy $M(n+1) = M_{M(n)} = 2^{M(n)} - 1$ starting from $M(1) = 2$ and giving Mersenne primes $3, 7, 127, 2^{127} - 1$. $M_7 = 127$ corresponds to genetic code. This inspires the idea that perhaps $(r_0 = M_7, s_0 = 1)$ might be worth of checking. The parameter values $r_0 = 127, s_0 = 1$ indeed yield the first example for which the spectral power for primes $p \leq 11$ is reasonably small and equal to 17. 22 units of spectral power however concentrates on $p = 2^5 - 1 = 31$, the Mersenne prime below M_7! many primes are lacking from the spectrum. In any case, it would seem possible to distribute the spectral power outside the small prime region but it is clear that genetic code would be number theoretically something extremely special of realized in this manner.

For $s_0 > 1$ spectral power again concentrates on $p = 2$. Since M_{127} corresponds to the Mersenne assigned to the memetic code, natural curiousity leads to check what happens in this case. All spectral power concentrates to $p = 2$ in this case: this is nothing but 2-adic spontaneous magnetization! It seems that this phenomenon occurs quite generally for very large values of r_0.

There might be something wrong with the program making the modulo arithmetics. For even values of r_0 partition function should be odd and $p = 2$ would give positive contribution to entropy. The general finding is that $p = 2$ is highly degenerate. This is possible only if the partition function fails to be divisible for primes $2 \leq p \leq 61$ for very many values of n. Even this does not help for $r_0 = 2^n$ since in this case $p > 2$ gives non-positive entropy for all values of n.

7 Appendix

The appendix sums up some computational aspects of the model and represents the models for doublet and singlet genetic codes as toy models.

7.1 Computational Aspects

7.1.1 Calculation of partition numbers $d(n, r)$

The basic problem in the calculation of partition numbers $p(n, r)$ is the presence of partitions containing same integer several times. This problem can be circumvent by arranging the integers in the partition in decreasing order so that one has $n_1 \geq n_2 ... \geq n_r$. Using this ordering the calculation of partition numbers $d(n, r)$

$$d(n,r) \quad = \quad \sum_{k=1}^{n-r+1} d(n-k, r-1|k) \ , \tag{7.1}$$

where $d(n, r|k)$ denotes the number of partitions for which the first number n_1 satisfies $n_1 \leq k$. The formula states that the ordered r-partitions of n decompose as $(k, n_1, ...n_{r-1})$, $k \leq n - r - 1$ such that $r - 1$-partition $(n_1, ..., n_{r-1})$ satisfies $n_1 \leq k$ by the ordering assumption.

What one must calculate are the numbers $d(n - k, r|k)$ and this can be done recursively

$$d(n,r|k) \quad = \quad \sum_{k_1 \leq k} d(n-k_1, r-1|k_1) \ . \tag{7.2}$$

The basic data item besides these formulas is $d(1, 1) = 1$. Also $d(n, n) = 1$ and $d(n, 1) = 1$ can be used.

The algorithm becomes time consuming for $n > 50$ and larger partition numbers are conveniently calculated by using the recurrence relation [A1]

$$P(n,k) \quad = \quad P(n-1, k-1) + P(n-k, k) \ . \tag{7.3}$$

The numbers $Q(n, k)$ of partitions of n to integers such that same integer does not appear twice are obtained from the formula [A1]

$$Q(n,k) \quad = \quad P(n - \begin{pmatrix} k \\ 2 \end{pmatrix}, k) \quad . \tag{7.4}$$

7.1.2 Numerical treatment of $n_0 < 0$ polynomial thermodynamics

The numerical treatment of $n_0 < 0$ polynomial thermodynamics is somewhat tricky and deserves a separate discussion. For definiteness the consideration is restricted to $H = log(r + r_0)$ case with $T = 1/n_0$. The generalization to other critical Hamiltonians is trivial.

For $n_0 = -m < 0$ case the entropy has the expression

$$
\begin{aligned}
S_p(n) &= \sum_{r=1}^{n} p(n,r)\left[mlog(|\frac{(n+r_0)}{r_0!(r+r_0)}|_p)\right] - log(|Z(n)|_p) \\
&= \left[\sum_{i=r_0+2}^{n+r_0} k_p(i) - \sum_r p(n,r)k_p(r+r_0)\right] mlog(p) - log(|Z(n)|_p) \\
Z(n) &= \sum_{r=1}^{n}\left[\frac{(n+r_0)!}{r_0!(r+r_0)}\right]^m d(n,r) \quad . \qquad (7.5)
\end{aligned}
$$

Here $k_p(n)$ is defined by the p-adic norm $|n|_p = p^{k_p}$. The integers appearing as coefficients of $d(n,r)$ in Z are very large and this causes numerical difficulties since factorials are represented precisely as integers only up to 21! and mod operation gives zero above this limit.

In order to calculate the p-adic norm of Z one must perform modulo p^k operations for Z by doing it separately for each summand and summing the resulting expressions. The problem is that the modulo p^k operation for the products involved does not reduce it to a small integer when p is large and one is forced to do the sum of large integers.

The solution of the problem is provided by finite field arithmetics. Start with the expression of $Z(n)$ written as

$$
Z(n) = \sum_{r=1}^{n} \frac{1}{(r+r_0)^m} d(n,r) \quad . \qquad (7.6)
$$

Since the calculation of p-adic norm involves only repeated modulo p operations to check whether the result vanishes modulo p, and if it does, a subsequent division by p, it suffices to interpret the factors $(1/(r+r_0)^m$ as elements of finite field $G(p,1)$.

1. If the condition $r + r_0 \; mod \; p \neq 0$ holds true, all denominators are non-vanishing. This is the case when $r_0 + 1 \leq p \leq n + r_0$ holds true. In this case it suffices to calculate the inverses $(r+r_0)_p^{-1}$ of $r + r_0$ in $G(p,1)$ and replace $Z(n)$ with

$$
\hat{Z}(n) = \sum_{r=1}^{n} \left[(r+r_0)_p^{-1}\right]^m d(n,r) \quad . \qquad (7.7)
$$

   The resulting expression is free of overflow problems and its p-adic norm can be calculated without difficulties.

2. When the condition $r + r_0 \; mod \; p \neq 0$ fails to be satisfied poles appear at $r = r_k = kp - r_0$, $k_{min} = [(1+r_0)/p] + 1 \leq k \leq k_{max} = [(n+r_0)/p]$, where $[x]$ denotes nearest integer smaller than x. Note that the problem is not encountered for $r_0 > 60$. The trick is to express Z in the form

$$
\begin{aligned}
Z(n) &= \frac{1}{X} \times \hat{Z}(n) \quad , \\
\hat{Z}(n) &= \sum_{r \neq kp - r_0} X \times \left[(r+r_0)_p^{-1}\right]^m \times d(n,r) \\
&+ \sum_{k=k_{min}}^{k_{max}} X_k \times d(n, kp - r_0) \quad , \\
X &= \prod_k (r_k + r_0)^m = \prod_{i=k_{min}}^{k_{max}} (ip)^m = \left(\prod_{i=k_{min}}^{k_{max}} i\right)^m p^{m(k_{max}-k_{min})} \quad , \\
X_k &= \frac{X}{(r_k+r_0)^m} = \left(\frac{\prod_{i=k_{min}}^{k_{max}} i}{k}\right)^m \times p^{m(k_{max}-k_{min}-1)} \quad . \qquad (7.8)
\end{aligned}
$$

This expression involves only relatively small integers and overflow problems are avoided.

$k_p(X)$ can be expressed in the form

$$k_p(X) = m\left[\sum_{k_{min}}^{k_{max}} k_p(k) - k_{max} + k_{min}\right] \ . \tag{7.9}$$

To sum up, the expression for $S_p(n)$ reduces in $(n_0 = -m < 0, r_0)$ case to the form

$$\begin{aligned}
\frac{S_p(n)}{log(p)} &= \left[\sum_{i=r_0+2}^{n+r_0} k_p(i) + \sum_{k_{min}}^{k_{max}} k_p(k) - k_{max} + k_{min} - \sum_{r=1}^{n} p(n,r)k_p(r+r_0)\right] m \\
&- k_p(\hat{Z}(n)) \ , \\
\hat{Z}(n) &= \sum_{r \neq kp - r_0} X\left[(r+r_0)_p^{-1}\right]^m d(n,r) + \sum_{k=k_{min}}^{k_{max}} X_k \times d(n, kp - r_0) \ , \\
X &= \left(\prod_{i=k_{min}}^{k_{max}} i\right)^m p^{m(k_{max}-k_{min})} \ , \\
X_k &= = \left(\frac{\prod_{i=k_{min}}^{k_{max}} i}{k}\right)^m \times p^{m(k_{max}-k_{min}-1)} \ , \\
k_{min} &= [(1+r_0)/p] + 1 \ , \ k_{max} = [(n+r_0)/p] \ .
\end{aligned} \tag{7.10}$$

In the nonsingular case 1) $X = 1$ and $X_i = 0$ holds true.

In practice $r \neq r_k$ terms do not contribute to $k(\hat{Z}(n))$ unless all $d(n, kp - r_0)$ happen to be divisible by a large power of p. The highest power of $p \leq 61$ appearing in $d(n,r)$ is 4 for $n \leq 63$. For the sake of generality and safety it is however better to keep also these contributions in the formula.

7.2 Number Theoretic Model For Singlet And Doublet Codes As AToy Model

The model of the genetic code applies to any number n of DNAs and maps the numbers $n = 0, 1...n-1$ to $\{0,1\} \cup \{primes\ p \leq n-1\}$. In [**?**] a model for the genetic code resulting via a symmetry breaking from the product of codes associated with 16 DNA doublets and 4 DNA singlets was considered. At the level of DNAs the product code is very natural and the almost symmetries of the genetic code with respect to last codon support the idea.

7.2.1 Singlet code

In the case of singlet code the requirement that at least single stopping sign codon exists, implies that either $p = 2$ or $p = 3$ fails to be coded. This would conform with the idea that $n = 3 = -1\ mod\ 4$ represents automatically stopping sign and 3 amino-acids would be coded. Fermionic entropy vanishes identically with this assumption.

It is perhaps instructive to consider the singlet codes at low temperature limit of exponential thermodynamics for $(r_0 > 1, s_0 = 1)$ to get some grasp of the situation. Singlet code gives $(\hat{d}(1), \hat{d}(2), \hat{d}(3)) = (1, 1+r_0, 1+r_0+r_0^2)$. The probabilities $p(n,r)$ are $p(n,r) = r_0^{n-r}/\hat{d}(n)$ and entropy can be written as

$$S_p(n) = -\frac{r_0^m}{1+r_0+..+r_0^n}\sum_{m=1}^{n} log(|\frac{r_0^m}{1+r_0+...+r_0^n}|_p) \ . \tag{7.11}$$

For $r_0 = 2$ *resp.* $r_0 = 3$ one has $(\hat{d}(1), \hat{d}(2), \hat{d}(3)) = (1,3,7)$ and $(\hat{d}(1), \hat{d}(2), \hat{d}(3)) = (1,4,13)$. For $r_0 = 2$ the code is $(0,1,2,3) \to (0,1,3,stop)$ with $n = 3$ having vanishing entropy and thus

naturally acting as stopping codon. $p = 2$ is not coded. For $r_0 = 3$ the code is $(0, 1, 2, 3) \to (0, 1, 2, stop)$. $p = 3$ is not coded.

Allowing $s_0 > 1$ does not allow to circumvent these problems. In this case the formula for entropy reads as

$$S_p(n) = -\frac{1}{s_0^n + r_0 s_0^{n-1} .. + r_0^n} \sum_{m=1}^{n} r_0^m s_0^{n-m} log(|\frac{r_0^m s_0^{n-m}}{s_0^n + r_0 s_0^{n-1} .. + r_0^n}|_p) \ . \tag{7.12}$$

For $(r_0 = 3, s_0 = 2)$ the denominator is not divisible by 2 or 3 so that all codons possess vanishing or negative information. The conclusion is that the mapping of $3 = -1 mod 4$ to stopping codon is the only consistent option.

For polynomial thermodynamics with Boltzmann weights given by $(r + r_0)^{n_0}$ there is a large number of parameter combinations giving single stopping codon which is always $n = 2$.

7.2.2 Doublet codes

Doublet code should map the integers $0, 1..., 14(15)$ to primes $0, 1, 2, 3, 5, 7, 11, 13$. The inspection of the tables 1 and 2 shows that at infinite temperature limit $p = 13$ fails to coded for both B, F, and BF and also $p = 7$ for F. $n = 13$ is not coded to a unique prime for B. The parameter values are restricted to the range $(n_0, r_0) \in (\{1, 5\}, \{0, 5\})$ in the polynomial case and to the range $(r_0, s_0)(\{1, 5\}, \{1, 5\})$ in the exponential case. The findings support the view that polynomial thermodynamics is the only viable approach.

1. Stopping sign codons as codons with $S_p < 0$

For finite temperature thermodynamics the conditions used are that least one stopping codon having by definition $S_p < 0$ exists and all primes $p \leq 13$ must be coded.

1. For finite temperature polynomial thermodynamics the cases F and BF allow no solutions whereas B allows four solutions $((n_0, r_0) = (1, 5), (2, 1),$ $(2, 5), (3, 3))$.

2. For exponential thermodynamics neither, B, F, nor BF allow solutions.

2. Stopping sign codons as $n = 15$ codon or codons with $S_p < 0$

One could argue that since $n = 15$ corresponds to -1 in modulo 16 mathematics, it should code for stopping sign. If so, the situation changes.

1. Polynomial thermodynamics.

   In BF case $(n_0, r_0) = (1, 1)$ provides in the range $(n_0, r_0) \in (\{1, 5\}, \{0, 5\})$ the only example of a genetic code for which all primes $p \leq 13$ are coded. One can say that supersymmetric option fixes the code uniquely in this parameter range. F allows no solutions. B allows 4 solutions $((n_0, r_0) = (1, 2), (2, 1), (2, 5), (3, 3))$.

2. Exponential thermodynamics

   Neither B, F, nor BF type thermodynamics allow solutions.

8 Galois groups and genes

In an article discussing a TGD inspired model for possible variations of G_{eff} [**?**], I ended up with an old idea that subgroups of Galois group could be analogous to conserved genes in that they could be conserved in number theoretic evolution. In small variations such as above variation Galois subgroups as genes would change only a little bit. For instance, the dimension of Galois subgroup would change.

n	$d_B(n)$	$p_B(n)$	$d_F(n)$	$p_F(n)$	$p_{BF}(n)$
0	1	1	1	0	0
1	1	1	1	1	1
2	2	2	1	1	2
3	3	3	2	2	3
4	5	5	2	2	5
5	7	7	3	3	7
6	11	11	4	2	11
7	3×5	5	5	5	5
8	2×11	11	6	3	11
9	$2 \times 3 \times 5$	5	8	2	2
10	$2 \times 3 \times 7$	7	10	5	3
11	$2^3 \times 7$	2	12	2	2
12	7×11	11	15	5	11
13	101 (prime)	?	18	3	3
14	$3^3 \times 5$	3	22	11	3
15	$2^4 \times 11$	2	27	3	3

Table 15: Table represents the partition numbers $d_B(n)$ and $d_F(n)$ as well as the primes $p_B(n)$, $p_F(n)$, $p_{BF}(n)$ resulting from the minimization of the p-adic entropy $S_{I,p}(n)$, $I = B, F, BF$ as a function of n for $n < 16$.

The analogy between subgoups of Galois groups and genes goes also in other direction. I have proposed long time ago that genes (or maybe even DNA codons) could be labelled by $h_{eff}/h = n$. This would mean that genes (or even codons) are labelled by a Galois group of Galois extension (see `http://tinyurl.com/zu5ey96`) of rationals with dimension n defining the number of sheets of space-time surface as covering space. This could give a concrete dynamical and geometric meaning for the notion of gene and it might be possible some day to understand why given gene correlates with particular function. This is of course one of the big problems of biology.

8.1 Could DNA sequence define an inclusion hierarchy of Galois extensions?

One should have some kind of procedure giving rise to hierarchies of Galois groups assignable to genes. One would also like to assign to letter, codon and gene and extension of rationals and its Galois group. The natural starting point would be a sequence of so called intermediate Galois extensions E^H leading from rationals or some extension K of rationals to the final extension E. Galois extension has the property that if a polynomial with coefficients in K has single root in E, also other roots are in E meaning that the polynomial with coefficients K factorizes into a product of linear polynomials. For Galois extensions the defining polynomials are irreducible so that they do not reduce to a product of polynomials.

Any sub-group $H \subset Gal(E/K)$) leaves the intermediate extension E^H invariant in element-wise manner as a sub-field of E (see `http://tinyurl.com/y958drcy`). Any subgroup $H \subset Gal(E/K)$) defines an intermediate extension E^H and subgroup $H_1 \subset H_2 \subset ...$ define a hierarchy of extensions $E^{H_1} > E^{H_2} > E^{H_3}...$ with decreasing dimension. The subgroups H are normal - in other words $Gal(E)$ leaves them invariant and $Gal(E)/H$ is group. The order $|H|$ is the dimension of E as an extension of E^H. This is a highly non-trivial piece of information. The dimension of E factorizes to a product $\prod_i |H_i|$ of dimensions for a sequence of groups H_i.

Could a sequence of DNA letters/codons somehow define a sequence of extensions? Could one assign to a given letter/codon a definite group H_i so that a sequence of letters/codons would correspond a product of some kind for these groups or should one be satisfied only with the assignment of a standard kind of extension to a letter/codon?

Irreducible polynomials define Galois extensions and one should understand what happens to an irreducible polynomial of an extension E^H in a further extension to E. The degree of E^H increases by a factor, which is dimension of E/E^H and also the dimension of H. Is there a standard manner

to construct irreducible extensions of this kind?

1. What comes into mathematically uneducated mind of physicist is the functional decomposition $P^{m+n}(x) = P^m(P^n(x))$ of polynomials assignable to sub-units (letters/codons/genes) with coefficients in K for a algebraic counterpart for the product of sub-units. $P^m(P^n(x))$ would be a polynomial of degree $n+m$ in K and polynomial of degree m in E^H and one could assign to a given gene a fixed polynomial obtained as an iterated function composition. Intuitively it seems clear that in the generic case $P^m(P^n(x))$ does not decompose to a product of lower order polynomials. One could use also polynomials assignable to codons or letters as basic units. Also polynomials of genes could be fused in the same manner.

2. If this indeed gives a Galois extension, the dimension m of the intermediate extension should be same as the order of its Galois group. Composition would be non-commutative but associative as the physical picture demands. The longer the gene, the higher the algebraic complexity would be. Could functional decomposition define the rule for who extensions and Galois groups correspond to genes? Very naïvely, functional decomposition in mathematical sense would correspond to composition of functions in biological sense.

3. This picture would conform with $M^8 - M^4 \times CP_2$ correspondence [**?**] in which the construction of space-time surface at level of M^8 reduces to the construction of zero loci of polynomials of octonions, with rational coefficients. DNA letters, codons, and genes would correspond to polynomials of this kind.

8.2 Could one say anything about the Galois groups of DNA letters?

A fascinating possibility is that this picture could allow to say something non-trivial about the Galois groups of DNA letters.

1. Since $n = h_{eff}/h$ serves as a kind of quantum IQ, and since molecular structures consisting of large number of particles are very complex, one could argue that n for DNA or its dark variant realized as dark proton sequences can be rather large and depend on the evolutionary level of organism and even the type of cell (neuron viz. soma cell). On the other, hand one could argue that in some sense DNA, which is often thought as information processor, could be analogous to an integrable quantum field theory and be solvable in some sense. Notice also that one can start from a background defined by given extension K of rationals and consider polynomials with coefficients in K. Under some conditions situation could be like that for rationals.

2. The simplest guess would be that the 4 DNA letters correspond to 4 non-trivial finite groups with smaller possible orders: the cyclic groups Z_2, Z_3 with orders 2 and 3 plus 2 finite groups of order 4 (see the table of finite groups in `http://tinyurl.com/j8d5uyh`). The groups of order 4 are cyclic group $Z_4 = Z_2 \times Z_2$ and Klein group $Z_2 \oplus Z_2$ acting as a symmetry group of rectangle that is not square - its elements have square equal to unit element. All these 4 groups are Abelian. Polynomial equations of degree not larger than 4 can be solved exactly in the sense that one can write their roots in terms of radicals.

3. Could there exist some kind of connection between the number 4 of DNA letters and 4 polynomials of degree less than 5 for whose roots one an write closed expressions in terms of radicals as Galois found? Could it be that the polynomials obtained by a a repeated functional composition of the polynomials of DNA letters have also this solvability property?

   This could be the case! Galois theory states that the roots of polynomial are solvable by radicals if and only if the Galois group is solvable meaning that it can be constructed from abelian groups using Abelian extensions (see `https://cutt.ly/4RuXmGo`).

   Solvability translates to a statement that the group allows so called sub-normal series $1 < G_0 < G_1 ... < G_k$ such that G_{j-1} is normal subgroup of G_j and G_j/G_{j-1} is an abelian group. An equivalent condition is that the derived series $G \triangleright G^{(1)} \triangleright G^{(2)} \triangleright ...$ in which $j+1$:th group is commutator group of G_j ends to trivial group. If one constructs the iterated polynomials by using only the 4 polynomials with Abelian Galois groups, the intuition of physicist suggests

that the solvability condition is guaranteed! Wikipedia article also informs that for finite groups solvable group is a group whose composition series has only factors which are cyclic groups of prime order.

Abelian groups are trivially solvable, nilpotent groups are solvable, p-groups (having order, which is power prime) are solvable and all finite p-groups are nilpotent. Every group with order less than 60 elements is solvable. Fourth order polynomials can have at most S_4 with 24 elements as Galois groups and are thus solvable. Fifth order polynomials can have the smallest non-solvable group, which is alternating group A_5 with 60 elements as Galois group and in this case are not solvable. S_n is not solvable for $n > 4$ and by the finding that S_n as Galois group is favored by its special properties (see `https://arxiv.org/pdf/1511.06446.pdf`).

A_5 acts as the group icosahedral orientation preserving isometries (rotations). Icosahedron and tetrahedron glued to it along one triangular face play a key role in TGD inspired model of bio-harmony and of genetic code [**?**, **?**]. The gluing of tetrahedron increases the number of codons from 60 to 64. The gluing of tetrahedron to icosahedron also reduces the order of isometry group to the rotations leaving the common face fixed and makes it solvable: could this explain why the ugly looking gluing of tetrahedron to icosahedron is needed? Could the smallest solvable groups and smallest non-solvable group be crucial for understanding the number theory of the genetic code.

An interesting question inspired by $M^8 - H$-duality [**?**] is whether the solvability could be posed on octonionic polynomials as a condition guaranteeing that TGD is integrable theory in number theoretical sense or perhaps following from the conditions posed on the octonionic polynomials. Space-time surfaces in M^8 would correspond to zero loci of real/imaginary parts (in quaternionic sense) for octonionic polynomials obtained from rational polynomials by analytic continuation. Could solvability relate to the condition guaranteeing M^8 duality boiling down to the condition that the tangent spaces of space-time surface are labelled by points of CP_2. This requires that tangent or normal space is associative (quaternionic) and that it contains fixed complex sub-space of octonions or perhaps more generally, there exists an integrable distribution of complex subspaces of octonions defining an analog of string world sheet.

What could the interpretation for the events in which the dimension of the extension of rationals increases? Galois extension is extensions of an extension with relative Galois group $Gal(rel) = Gal(new)/Gal(old)$. Here $Gal(old)$ is a normal subgroup of $Gal(new)$. A highly attractive possibility is that evolutionary sequences quite generally (not only in biology) correspond to this kind of sequences of Galois extensions. The relative Galois groups in the sequence would be analogous to conserved genes, and genes could indeed correspond to Galois groups [K2] [**?**]. To my best understanding this corresponds to a situation in which the new polynomial P_{m+n} defining the new extension is a polynomial P_m having as argument the old polynomial $P_n(x)$: $P_{m+n}(x) = P_m(P_n(x))$.

What about the interpretation at the level of conscious experience? A possible interpretation is that the quantum jump leading to an extension of an extension corresponds to an emergence of a reflective level of consciousness giving rise to a conscious experience about experience. The abstraction level of the system becomes higher as is natural since number theoretic evolution as an increase of algebraic complexity is in question.

This picture could have a counterpart also in terms of the hierarchy of inclusions of hyperfinite factors of type II_1 (HFFs). The included factor M and including factor N would correspond to extensions of rationals labelled by Galois groups $Gal(M)$ and $Gal(N)$ having $Gal(M) \subset Gal(M)$ as normal subgroup so that the factor group $Gal(N)/Gal(M)$ would be the relative Galois group for the larger extension as extension of the smaller extension. I have indeed proposed [**?**] that the inclusions for which included and including factor consist of operators which are invariant under discrete subgroup of $SU(2)$ generalizes so that all Galois groups are possible. One would have Galois confinement analogous to color confinement: the operators generating physical states could have Galois quantum numbers but the physical states would be Galois singlets.

REFERENCES

Mathematics

[A1] Partition function P. Available at: https://mathworld.wolfram.com/PartitionFunctionP.html.

[A2] Frenkel E. Representation Theory: Its rise and Its Role in Number Theory, 2004. Available at: https://www.sunsite.ubc.ca/DigitalMathArchive/Langlands/pdf/gibbs-ps.pdf.

[A3] Mahlburg K. Partition congruences and the andrews-garvan-dyson crank. Available at: https://www.jstor.org/pss/4143442.

Biology

[I1] The Genetic Code. Available at: https://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Codons.html.

[I2] Horgan J. The World According to RNA. *Sci Am*, 1996.

[I3] Gilbert W. The RNA World. *Nature*, 319, 1986.

Books related to TGD

[K1] Pitkänen M. Topological Quantum Computation in TGD Universe. In *Quantum - and Classical Computation in TGD Universe. https://tgdtheory.fi/tgdhtml/Btgdcomp.html*. Available at: https://tgdtheory.fi/pdfpool/tqc.pdf, 2015.

[K2] Pitkänen M. Could Genetic Code Be Understood Number Theoretically? In *Genes and Memes: Part I. https://tgdtheory.fi/tgdhtml/genememe1.html*. Available at: https://tgdtheory.fi/pdfpool/genenumber.pdf, 2023.

[K3] Pitkänen M. Crop Circles and Life at Parallel Space-Time Sheets. In *Magnetospheric Consciousness. https://tgdtheory.fi/tgdhtml/Bmagnconsc.html*. Available at: https://tgdtheory.fi/pdfpool/crop1.pdf, 2023.

[K4] Pitkänen M. Crop Circles and Life at Parallel Space-Time Sheets. In *Magnetospheric Consciousness. https://tgdtheory.fi/tgdhtml/Bmagnconsc.html*. Available at: https://tgdtheory.fi/pdfpool/crop2.pdf, 2023.

[K5] Pitkänen M. Many-Sheeted DNA. In *Genes and Memes: Part I. https://tgdtheory.fi/tgdhtml/Bgenememe1.html*. Available at: https://tgdtheory.fi/pdfpool/genecodec.pdf, 2023.

[K6] Pitkänen M. Negentropy Maximization Principle. In *TGD Inspired Theory of Consciousness: Part I. https://tgdtheory.fi/tgdhtml/Btgdconsc1.html*. Available at: https://tgdtheory.fi/pdfpool/nmpc.pdf, 2023.