Genes and Memes

M. Pitkänen,

February 16, 2022
Email: matpitka6@gmail.com.
http://tgdtheory.com/public_html/.
Recent postal address: Rinnekatu 2-4 A 8, 03620, Karkkila, Finland.

Contents

Abstract

In this article basic TGD inspired ideas about genetic code are discussed.

1. Genetic and memetic code from the model of abstraction process

The basic numbers of genetic code are probably not accidental. This led for more than two decades ago to an attempt to construct a model for abstraction process reproducing the basic numbers of the genetic code. The simplest model for an abstraction process is based on a repeated formation of statements about statements starting from two basic statements. If one drops at each step of the construction the statement corresponding to empty set in the set theoretic realization of Boolean algebra, one obtains a hierarchy allowing to understand the basic numbers of genetic code, including the number of amino-acids. What one obtains is so called Combinatorial Hierarchy consisting of the Mersenne numbers $2, M(1) = 3, 7$,$127, 2^{127} - 1, ..$ constructed using the rule $M(n+1) = M_{M(n)} = 2^{M(n)} - 1$. The explicitly listed ones are known to be primes. Combinatorial Hierarchy emerges from a model of abstraction process as subsequent transitions from level to meta level by forming Boolean statements about Boolean statements of level n and dropping one statement away.

The infinite hierarchy of possible genetic codes suggests the possibility of an infinite hierarchy of increasingly complex life-forms. The natural question is whether a counterpart of the genetic code could make sense for our ideas, memes. Combinatorial Hierarchy model for abstraction process predicts that memetic code should correspond to the level M_{127} of the hierarchy. This leads to a precise realization of the memetic code in terms of binary sequences. Codewords, counterparts of mRNA, correspond to 126-bit sequences. Also almost-127-bit code with $2^{127} - 1$ codons is possible.

2. Frequency and pulse representations of codes

p-Adic length scale hypothesis and identification of codes as special cases of a hierarchy of p-adic cognitive codes allows quantitative predictions. The most general assumption assigns to any prime $p \simeq 2^k$, k integer, a hierarchy of cognitive codes with codeword having a duration equal to n-ary p-adic time scale $T_p(n)$ such that the number of bits is factor k_1 of k. Codewords could be realized either as k_1 harmonics of the fundamental frequency $f_p(n) = 1/T_p(n)$ or as temporal sequences of bits of duration $\tau = T_p(n)/k_1$ represented as pulses of maximal duration τ. Pulse-frequency dichotomy corresponds to dichotomies like particle-wave, nerve pulse-EEG, and talking left brain-singing right brain.

Genetic code would correspond to $k = 2^7 - 1 = 127$ and have 6 bits (64 DNA triplets). These codewords could be realized dynamically as temporal field patterns. For genetic code primes $p \simeq 2^k$, $k = 6 \times n$ define candidates for the duration of the genetic code word if all factors of k are assumed to define a possible number of bits of the code word. The time scales come as powers of 8 so that they cover the entire range of biologically relevant time scales down to CP_2 length scale, and genetic code could appear as fractally scaled versions unlike memetic code and perhaps also outside the biological context. $k = 2 \times 126 = 2 \times 6 \times 21 = 252$ allows the representation of both 126-bit memetic codeword, 6-bit genetic codeword, and almost-7-bit genetic code word. For pulse representation genetic codon would have a duration of 50 ms whereas the bit would have duration of 8.3 ms so that the realization using nerve pulse patterns is in principle possible. Frequency representation would be realized as 6 first harmonics of the fundamental frequency $f_1 = 2^n \times 20$ Hz, where $f_1 = 20$ Hz defines the lower end of audible frequency range and also the rate for the translation of mRNA triplets to amino-acids. 126-bit memetic code allows a representation as sequence of 21 nerve pulses of duration 2.4 ms each of them accompanied by 6-bit genetic codon realized at the microtubular level (this representation of genetic code has been suggested by Koruga).

The secondary p-adic time scale associated with M_{127} is .1 seconds and defines the duration of the almost 127-bit memetic codeword. For frequency representation is realized as 127 first harmonics of $f_1 = 10$ Hz and the duration of the bit for pulse representation is .8 ms which is shorter than the duration of nerve pulse. The duration .1 seconds of code word might be identified as the minimal duration of cortical mental images, and the so called features introduced by Walter Freeman could define pulse representation of memetic code words of 127 bits. The highest frequency in the frequency representation is 1270 Hz and could define the frequency responsible for synchronous neuronal firing known to be about 1 kHz. Various numerical co-incidences suggest that language corresponds to a particular realization of memetic and genetic codes closely related to their realization at DNA level.

3. Model for the evolution of genetic code from the symmetries of the code

TGD leads to a model for the evolution of the genetic code motivated by the observation that the genetic code possesses an exact A-G and almost exact T-C permutation symmetry with respect to the third nucleotide of the DNA triplet. This leads to the hypothesis that genetic code has evolved as a fusion of doublet and singlet codes accompanied by a small

breaking of the product symmetry. The hypothesis is highly predictive, and it is possible to reproduce genetic code and its variants by this mechanism in a natural manner. The mechanism has deep implications for the models of the bio-chemical evolution before genetic code: in particular a detailed model for the evolution of genetic code and pre-biotic evolution emerges.

4. Mapping memetic code to 169-bit micro-tubular code

169-bit micro-tubular code words is excellent candidate for a representation of long term memories as a temporal list of activated memes. The model for the mapping of memetic code to 169-bit microtubular code is dictated by the general ideas about realization of intentions and p-adic cognitive codes. When combined with general number theoretical arguments and physical considerations the model becomes highly unique. The prediction for the intronic representation of the memetic codon involving 9 DNA triplets as parity bits is readily testable, and also the prediction for the microtubular electric field pattern is in principle testable.

5. Genes, memes, and universal language

Also static representations of the memetic code are possible and intronic DNA could provide representation of memetic codewords as sequences of 21 DNA triplets. At DNA level memes and genes should relate like computer software and hardware. In the case of language the rules producing a given linguistic expression can be seen as the high level software, main programs, whereas words can be seen as hardware-like lower level subprograms. This leads to the idea that memetic codewords define the basic program modules producing linguistic expressions by activating genes which express themselves in terms of field patterns generating nerv pulse patterns generating words or word sequences very much analogous to proteins.

Time mirror mechanism and the structure of the computer language LISP inspire a concrete model for memes as intronic programs initiated from magnetic body and calling genes as subprograms in turn calling other genes as subprograms and generating at the lowest level field patterns generating nerve pulses patterns giving rise to the motor action producing speech. Phonemes could directly correspond to DNA triplets and define the basic building blocks of language having as such no meaning. If this view is correct, the development of spoken and written language would mean basically the emergence of a higher level of intentionality, which utilizes an already existing repertoire of memes expressed in many other manners. This would in turn suggest that animals and even plants possess some kind of languages realized at cellular level, and that even inter-species communications using common memetic grammar and genetic vocabulary.

6. Corals and men

A strong support for the idea of interspecies communications come from the sensational finding that the genome of corals, known to be the most primitive animals having nervous system, share a large number of common genes with vertebrates whereas they share much less common genes with flies and worms. This finding challenges profoundly the existing view about the evolution of animals and adds a further mystery to the halo of mysteries surrounding Cambrian explosion.

Since corals are usually regarded as relatively simple creatures, the most obvious questions concern the function of the complex genome. The TGD inspired answer is that the common genes provide a common vocabulary making possible communications between corals and vertebrates such as fishes. The genes express themselves in terms of electromagnetic field patterns and cyclotron transitions of Ca_{++} ions giving rise to primitive EEG are crucially involved. The calcium containing skeleton possessed by both corals and vertebrates could amplify the field patterns representing genes and make possible interspecies communications.

Coral reefs can be also seen as super organisms with cells replaced by double cell layers forming the corals. This forces to consider the possibility that coral reefs are super-organisms perhaps even possessing super-neural system consisting of super-neurons defined by differentiated corals. Accordingly, in TGD Universe coral reefs could be seen as descendants of higher level intra-terrestrial life forms which boosted Cambrian explosion by horizontal transfer of genes to much simpler life forms and providing also them with a nervous system.

7. Does ontogeny recapitulate also the future phylogeny at the level of genes and memes?

Ontogeny recapitulates phylogeny means that the morphogenesis of the embryo repeats the evolutionary steps leading to the organism. One might ask whether and how this process is realized at the level of genes and memes (introns expressing themselves electromagnetically): this could provide further understanding of the mysterious "junk DNA". Combining this question with some recent puzzling findings leads to a rather radical revision of the view about evolution proceeding through random mutations.

1. The second strange finding besides coral genome reported in New Scientist (5 June,

2004) was that the removal of large portions of conserved intronic DNA from mice has no detectable effects on the basic biological functions. Conserved parts of DNA are usually thought as being an outcome of a long selection process and far from genetic trash. This could be understood if the conserved introns have been radiated from corals and the selection process has occurred already before the Cambrian explosion induced by the emergence of the corals and leading to the sudden emergence of new highly developed life forms. That mouse introns did not have any identifiable function could mean that they are still waiting for time to become ripe for their expression.

2. A third strange discovery relates to morphogenesis and is known as Ciba Geigy effect. Chemists Guido Ebner and Guido Schuerch exposed germs, seeds, and eggs to an electric field with strength in the range .5-2 kV/m. For instance, the resulting trouts appeared to resemble their ancient predecessors. The leaves of certain plants represented a series of snapshots from evolution with the oldest leaves dating back to 300 million years. This suggests that the memone and genome represent ontogeny recapitulates phylogeny principle quite concretely, and that static electric fields could provide the practical manner to activate and study the ancient morphologies. Even partial transmutation of life forms to each other might be possible (beautiful swan to ugly duckling at least!). The activation of morphologies not yet realized is probably more difficult: new memetic programs require new genetic hardware.

The resulting vision about evolution of higher organisms would be as the activation of conserved memes and genes basically inherited from corals rather than by the emergence of new genes by random mutations. Very much like learning new features of a text processing program. The explosive evolution of human civilization could correspond to a rapid shift of the activated portion of memone and genome. The fact that 95 per cent of our DNA consists of introns suggests that an enormous evolutionary potential exists also at the level of personal evolution during single life cycle. TGD view about space-time as a 4-dimensional living organism would mean that this personal evolution continues after the biological death since the 4-body of geometric past does not disappear in the biological death.

1 Introduction

In TGD based model of consciousness genes and memes are in very similar position. The quantum self-organization process explains genetic code as an outcome of Darwinian selection basically carried out by dissipation, and also memes should be subject to similar selection. Cognitive codes would be for neuroscience what genetic code is for biology. Hence there are good motivations to guess what the precise form and realization of these codes might be if it really exists. In fact, p-adic length scale hierarchy suggests an entire fractal hierarchy of cognitive codes with quantized durations of the codeword and preferred numbers of bits per code word [K14] . There exists however some preferred cognitive codes. A simple model for abstraction process leads to an entire hierarchy of genetic codes and one of them explains the basic numbers of the genetic code. A fascinating possibility is that the next level in the hierarchy yields memetic code and also this possibility will be studied in the sequel. Genetic and memetic codes represent particular examples associated with Mersenne primes $p = 127 = 2^7 - 1$ and $p = M_{127} = 2^{127} - 1$.

1.1 Combinatorial Hierarchy Of Codes

The basic numbers of genetic code are probably not accidental. This led for more than ten years ago to an attempt to construct a model for abstraction process reproducing the basic numbers of the genetic code.

1.1.1 Genetic code from a model for abstraction process

The simplest model for an abstraction process is based to a repeated formation of statements about statements starting from two basic statements representing the most primitive logical thoughts. If one drops at each step of construction the statement corresponding to empty set in the set theoretic realization of Boolean algebra, one obtains a hierarchy allowing to understand the basic numbers of genetic code.

The outcome is the so called Combinatorial Hierarchy [A3] consisting of the Mersenne numbers $2, M(1) = 3, 7$,$127, 2^{127} - 1, ..$ constructed using the rule $M(n+1) = M_{M(n)} = 2^{M(n)} - 1$. The explicitly listed ones are known to be primes. Combinatorial Hierarchy emerges from a model of abstraction process as sub-sequent transitions from level to meta level by forming Boolean statements about Boolean statements of level n and dropping one statement away. Combinatorial Hierarchy results also by constructing the sets of all subsets with empty set excluded starting from two element set. The set of statements at level n can be given a structure of Finite Field $G(M(n), 1)$ if $M(n)$ is prime. The multiplicative groups $Z_{M(n)-1}$ form a nested hierarchy and the coset spaces $Z_{k_n} \equiv Z_{M(n)-1}/Z_{M(n-1)-1}$ are cyclic groups ($k_n = (M(n) - 1)/(M(n-1) - 1)$). Hilbert's conjecture states that each Mersenne number in the Combinatorial Hierarchy is prime.

Combinatorial Hierarchy based model of genetic code explains the number of DNAs and amino-acids, and the representation of words of the genetic code as triplets of 4 different lower level codewords. Genetic code corresponds to $n = 3$ level of the hierarchy with 127 statements representable as 7-bit sequences with the sequence of seven "0":s dropped away. Only the 64 6-bit code words can be fully realized and correspond to $(M(3) + 1)/2 = 64$ DNA triplets. $k_3 = 126/6 = 21$ equals to the number of amino-acids plus stopping codon. There is a natural imbedding of subgroup Z_{21} identifiable as a representation of amino-acids to the group $Z_{126=6\times 21}$.

More abstractly, at level n the counterparts of DNA triplets correspond to the set $X_{N(DNA)} \subset Z_{M(n)-1}$ of $N(DNA) = (M(n) + 1)/2$ statements consistent with a fixed atomic statement (64 for $n = 3$). Atomic statement corresponds to a fixed value, assumed to be one, of a fixed bit in a bit sequence representation and a subset consisting of single element in the set theoretic representation. These statements could be regarded as statements consistent with the axiom defined by the selection of the atomic statement. The counterparts of amino-acids and stopping codon correspond to k_n theorems of a formal system defined by n^{th} level of Combinatorial Hierarchy having a unique imbedding as the group $Z_{k_n} \subset Z_{M(n)-1}$. The DNAs coding for a given "amino-acid" correspond to the special cases of the theorem.

Mapping of DNA code words to amino-acids generalizes to the mapping $x \to x^{k_{n-1}}$ in $Z_{M(n)-1}$ mapping DNA type statements to amino-acid type statements. $(M(n)+1)/2$ DNAs can be imbedded to Z_{126} with several manners. Genetic code is fixed ones this imbedding is given. For $n = 3$ one obtains ordinary genetic code defined by the map $x \to x^6$ and imbedding of the DNAs to Z_{126}. The numbers of DNA:s coding single amino-acid can be reproduced by a symmetry breaking mechanism involving the finite groups $Z_{p_{n-1}}$ and Z_{k_n} and symmetry breaking is in a well defined sense minimal. The infinite hierarchy of possible genetic codes (at least if Hilbert's conjecture holds true) suggests the possibility of an infinite hierarchy of increasingly complex life forms.

If one allows only Mersenne primes, the model for the abstraction process predicts at least one further code, which I have used to call memetic code. It corresponds to the Mersenne prime $M_{127} = 2^{127} - 1$ and has 2^{126} code words and $(2^{126} - 1)/(2^6 - 1)$ "amino-acids". The secondary p-adic time scale $T(2, M_{127})$ is .1 seconds and defines a fundamental time scale in bio-systems.

There are reasons to expect that memetic code is an especially interesting higher level cognitive code and realized in terms of field patterns. In particular intronic portion of DNA could realize memetic codewords as sequences of 21 DNA triplets and memes would define the counterparts of computer programs at DNA level whereas genes would define the counterpart of computer hardware coded into lower level programs and built only when needed. Both memes and genes could express themselves in terms of field patterns.

Memes and genes should relate like computer software and hardware. In the case of language the rules producing a given linguistic expression can be seen as the software whereas words can be seen as the hardware built from phonemes. This leads to the idea that memetic codewords define the basic program modules producing linguistic expressions by activating genes which express themselves as words or word sequences. Phonemes could directly correspond to DNA triplets and define the basic building blocks of language having as such no meaning. If this view is correct, the development of spoken and written language would mean basically the emergence of a higher level of intentionality, which utilizes an already existing repertoire of memes already expressed in many other manners. This would in turn suggest that animals and even plants possess some kind of languages realized at cellular level, and that even inter-species communications using common memetic vocabulary.

1.1.2 Myth about Fall of Man as metaphor for codes

There are basically two genetic and memetic codes: 6-bit and almost-7-bit ($2^3 - 1$ code words represented consciously) genetic codes and 126- and almost-127 bit memetic codes. The myth about the Fall of Man provides a little bit tongue-in-cheek metaphor for the extension of 126-bit code to almost-127-bit code. God gave to Adam and Eve the moral code as a single moral law formulated as an atomic statement "Do anything but do not eat from the Tree of Good and Bad Knowledge". Adam and Eve of course did this and probably as many fractally scaled versions.

The first Fall occurred probably already at the molecular level and meant the replacement of 6-bit genetic code with almost-7-bit code: $63 = 9 \times 7$ sins beside 64 good deeds appeared (there are 9 classes of sins containing one sin for every day of the week, one of them containing the seven deadly ones!). Imagination and intentions realized as deeds emerged and Adam and Eve became moral agents. For the memetic code the replacement of 126-bit memetic code led to almost 127-bit code and the repertoire of good and bad deeds was now much more impressive: 2^{126} good deeds and $2^{126} - 1$ sins (the basic classification sins to 7 basic types makes senses still!). One can of course wonder what was this un-doable sin with strong Gödelian flavor. A code word with only zeros cannot give rise to a conscious experience. Hence the un-doable sin could be such that the sinner cannot experience its consequences in this life, that is suicide.

1.1.3 Genetic code and thinking at DNA level

TGD predicts entire infinite hierarchy of selves starting from elementary particle level so that consciousness should be present also at DNA and protein level. The notion of self indeed allows to understand protein folding, which is rather mysterious phenomenon in standard physics framework.

The physical model of the genetic code constructed in [K8]. when interpreted in terms of the model for an abstraction process, suggests the interpretation of the genetic code as mapping the fundamental 64 truths to 20 basic conscious experiences, perhaps the protein level emotional experience about truth-ness: it must be emphasized that our experiences are probably not in question. Amino-acid P could correspond to the emotionally experienced truth "$G_1(P)$ or $G_2(P)$.. or $G_n(P)$ is true", where G_i code for protein P. 3 stopping sign codewords cannot be experienced emotionally as truths not non-truths (holy trinity at protein level!).

In the model for introns discussed in the chapter [K8] fermionic realization of Boolean statements plays an essential role in the genetic program of eukaryotes and suggests that Boolean mind appears already at molecular level and could correspond to the logical statements represented by genes. This Boolean mind does not correspond to our conscious logical thinking.

1.1.4 Does memetic code emerge at the next level of abstraction process?

The natural question is whether a counterpart of the genetic code could make sense for memes. Combinatorial Hierarchy model for abstraction process that memetic code should correspond to the level M_{127} of the hierarchy. This leads to a precise realization of the memetic code in terms of binary sequences. The secondary p-adic time scale associated with M_{127} is .1 seconds, which seems to define the duration for the immediate subjective memory. If this time scale corresponds to a sequence consisting 127 bits, the duration of single bit is 1/1270 seconds, which happens to be very near to definite p-adic time scale but is somewhat shorter than the typical duration of nerve pulse. This suggests that nerve pulse patterns as such cannot realize the full memetic code. The time scale associated with $k = 252 = 2 \times 6 \times 21$ is.05 seconds and one half of the duration of 127-bit memetic codon. $k = 252$ allows the representations of both 6- and 7-bit genetic codes and of the 126-bit memetic code.

An attractive hypothesis is that the temporal sequences for the Z^0 magnetization directions for a block cognitive antineutrinos at cell membrane space-time sheet provide a representation of the almost 127-bit memetic codeword. The conscious experience results when the Z^0 magnetization directions flip back to the direction of external Z^0 magnetic field in spin flipping cyclotron transition. $M_{127} = 2^{127} - 1$ different conscious experiences results since nothing happens if all cognitive antineutrinos are in the same direction as external Z^0 magnetic field. Z^0 magnetization direction could be altered by the Z^0 magnetic pulse associated with the Z^0 ME inducing cell membrane oscillations of nerve pulse pattern.

This raises obvious questions. Does genome have a memetic counterpart; what would be the function of the memone; what would be the memetic counterparts of the transcription and translation processes for genes? The natural guess is that memes are the basic building blocks of cognition and language. Perhaps also memes are coded by DNA, most naturally by introns, whose portion from the genome increases with the evolutionary level of organism and is 99 per cent for Homo Sapiens. The sequences of 21 DNA triplets would naturally realize 126-bit memetic codons 2^{126} memetic codons at DNA level. The prediction is that the intronic portions of the genome should consist of basic units containing 21 DNA triplets. This would also mean that the language conscious-to-us is only a tip of an iceberg. The intronic part of the DNA would be expressed in terms of MEs and involve communications between cell membrane and nucleus. The dynamics of this realization would be fast and nucleus would play the role of cellular brain.

1.2 The Product Model For The Evolution Of Genetic Code

It became as a surprise to me personally that the genetic code has an exact A-G permutation symmetry and an almost exact T-C permutation symmetry with respect to the third nucleotide. Seen with the eyes of a theoretical physicist knowing the enormous importance of spontaneous symmetry breaking in physics, these simple symmetries point the way to the understanding of the basic mechanism behind the evolution of the genetic code.

This inspired a simple model for our genetic code allowing to see the genetic code as a product of much simpler doublet and singlet codes with a small symmetry breaking due to the interaction between singlets and doublets. This model, even admittedly rather formal, might have deep implications for the theories how the life at the molecular level has involved. The physical realization of this model will be also discussed briefly. A detailed discussion can be found in [K19].

1.3 General Ideas About Codes And Languages

By quantum-classical correspondence space-time sheets provide a symbolic representation for the contents of consciousness. Therefore one can say that everything in principle represents and the task is to understand how these symbolic representations are generated, how codes are established, and how these symbolic representations generated the desired mental images. This obviously means a profound departure from the basic belief system of standard biology.

Computer languages form a hierarchy such that highest level languages are very flexible approaching gradually to the spoken language whereas lowest level languages are very precise and rigid. The notion of self hierarchy suggests that our spoken language is only a top of an iceberg and that below it is a hierarchy of languages ending down to the cellular level and DNA is one particular example about "computer language" realized in terms of p-adic cognitive codes, in particular genetic and memetic codes. In an attempt to understand whether and how memetic and other p-adic cognitive codes might relate to the spoken and written language one must have some general ideas codes and language.

1.3.1 The hierarchy of cognitive codes

p-Adic length scale hypothesis suggests an entire hierarchy of cognitive codes and languages. The primes $p \simeq 2^k$, k integer seems to be interesting physically, and prime values of k seem to be especially interesting. The codes would be characterized by the duration of the codeword given by n-ary p-adic time scale $T_p(n) = p^{(n-1)/2}T_p$, $T_p = 2^{k/2-127}T(2,127)$, $T(2,127) = .1$ seconds. The most general assumption is that number k_1 of bits of the codeword for given integer k_1 corresponds to some factor of k, the largest factor maximizing the information content. Codes could be represented either as temporal sequences of bits represented as pulses of maximal duration $T_p(n)/k_1$ or as superpositions of k_1 harmonics of $f_1 = 1/T_p(n)$, where Fourier components having intensity above/below critical value would represent bit 1/0. These representations will be referred to as pulse and frequency representations. Frequency representations would be realized in terms of topological light rays ("massless extremals", MEs) representing topologically quantized transverse radiation and pulse representations in terms of scalar wave pulses not possible in Maxwell's electrodynamics [K3]. This representational dichotomy reflects particle-wave duality and talking left brain and singing right brain dichotomy.

Memetic and genetic codes represent special examples of cognitive codes. One must distinguish between two representations: the representations involving 6 bits or almost 7 bits and 126 bits or almost 127 bits. "Almost" means that only $2^k - 1$ bit sequences rather than all 2^k bit sequences are realized as conscious bits if bits are realized as phase transitions.

Codes are always involved with classical communications involving transformation of mental images to a symbolic representation by some code. At our level of the hierarchy this symbolic representation could be speech, written language, picture, body language... This would suggest that also p-adic cognitive codes are involved with conscious communications. If these codes are realized in living systems, the bit sequences with the predicted durations and bit contents should induce biological effects serving as correlates for the conscious understanding of the message generated by the codewords at some level of the hierarchy.

TGD based view about living matter relies on the notion of field body or magnetic body associated with any system and having size much larger than the material body. Also these bodies form a fractal hierarchy. The communications from material body to field body could be based on cognitive codes. Given p-adic frequency corresponds f_p to a p-adic length scale $L_p = c/f_p$ characterizing the size of the magnetic body involved and for EEG frequencies the size scale of Earth is natural unit. For instance, p-adic cognitive codes realized in terms of field patterns would be involved with the communication of long term declarative memories from the geometric past.

1.3.2 What language is?

The attempt to understand the possible role of memetic code, a rough vision about what language is, allows to eliminate several ideas which look promising at first.

1. Language involves generation of symbolic representation of a mental image by a more or less rigid code. An example of a very flexible code is code based on associations. The symbolic representation of mental image should induce in the receiver the original mental image as faithfully as possible. This requires that a lot of common context. In particular, the neurologies and biologies of the sender and receiver must resemble each other sufficiently. In the case of high level languages like ordinary language even this is not enough and only simplest verbal signals and body language are understood universally. The cognitive codes associated with say cell level communications might make possible communications between cells of even different species remaining however unconscious to us.

2. The p-adic vision about evolution of cognitive skills like spoken language is that they evolve from long time and length scales to shorter ones. First a rough sketch about the motor action is created and gradually more and more details are added. This applies also at the level of the evolution of language itself. Simple signals expressing and generating emotions evolve gradually to spoken language which evolves to written language which in turn evolves to computer languages.

3. Learning of language requires learning of the conventions assigning to a given symbol a mental image. Sharing of mental images which represent more primitive "telepathic" communication makes possible this process. The observation that even plants and cells can react to our emotions and that this reaction does not depend much on distance [J5]. suggest that the sharing of mental images is in question. This allows to consider the possibility of inter-species linguistic communications using field patterns.

4. The understanding of language requires transformation of symbolic representation to conscious experience and here the notion of conscious bit ("cbit" [K11]) realized as a phase transition or as an absence of phase transition suggests itself. Phase transition could correspond to magnetization or formation of electret state and living matter could generate these representations in various length scales.

5. In TGD Universe intentions are realized as actions by a process, which proceeds from the magnetic body downwards along the hierarchy much like a desire of a boss of some institution to the lower levels of hierarchy. At each level intention or intentions are transformed to desires communicated to the lower levels of hierarchy. Intentions have p-adic space-time sheets as space-time correlates and are transformed to real ones representing the desire.

The most plausible realization of this process is in terms of time mirror mechanism. The space-time sheets in question would correspond to negative energy topological light rays representing the propagation of signals to the geometric past and induce processes. The process would continue down to the level of neurons and even DNA level and generate the desired action as a reaction to the resulting complex of desires. The beauty of the mechanism is that the communication to the geometric past makes it instantaneous.

Spoken and written language would rely on the same process and could propagate down to the level of genome and select the memes to be expressed. The expression of these memes as field patterns would then be a process propagating upwards in the hierarchy and finally generating speech or written word. When I decide to say something say the words "time mirror", this intention is transformed to a desire communicated to the geometric past to the lower level of the self hierarchy, and that at this level this desire generates further desires communicated to the lower levels. Ultimately this process ends down to the level of cells and even cell nuclei and DNA and induces response which propagates to the higher levels as neural and other activities inducing muscular activities in speech organs and generates the words "time mirror".

The signal to the geometric past involves negative energy photons and topological light rays. The working hypothesis has been that the signal to the geometric past is only a space-time correlate for sharing of the desire to generate the action, and does not involve any code. If this is the case then only the response propagating to the geometric future would be classical signal based on some code. One must however keep mind open to the possibility that also communications to the geometric past involve code.

1.3.3 Conscious bits and cognitive representations

The symbols representing message must be transformed to standardized mental images. The simplest possibility is that the mental images are coded to patterns of conscious bits or cbits. The general model for sensory and other qualia suggests that conscious bits should be realized as quantum jumps sequences associated with phase transitions. In this manner same quantum number increment is occurs for many particle for single quantum jump and for sufficiently long sequence of quantum jumps. Bit 1 would correspond to the occurrence of phase transition and bit 0 to the non-occurrence of the phase transition. For a code of k bits this has important implication: the codeword containing only zeros does not generate any conscious experience so that the number of experienced code words is $2^k - 1$. This could explain why Mersenne primes seem to be define especially important p-adic time scales.

Living matter is populated by dynamical electrets so that phase transitions between ordinary and electret states at various length scales are expected to be of special importance. Also magnetization of super phases at magnetic flux tubes of say Earth's magnetic field is expected to be one mechanism producing basic qualia serving as as bits.

1.3.4 Computer metaphor at DNA level

Software and hardware are essential elements of the computer and at DNA level this could mean that genes code for hardware which is not stable as in case of ordinary computers. This means that computer hardware is replaced by the possibility to generate it and genes carry the information needed for this. Introns would in turn represent the software, the programs and therefore also the linguistic aspect of DNA. An interesting possibility is that introns realize memes as sequences of 21 DNA triplets. This picture allows and even suggests that even DNA level might be involved with the generation of spoken words and define the deep structure of language.

The appendix of the book gives a summary about basic concepts of TGD with illustrations. Pdf representation of same files serving as a kind of glossary can be found at `http://tgdtheory.fi/tgdglossary.pdf` [?].

2 Combinatorial Hierarchy And Genetic Code

It is already found that Combinatorial Hierarchy emerges as a unique hierarchy of mappings of Boolean thoughts to association sequences, which could perhaps correspond to nerve pulse patterns,

or more probably, to temporal field patterns. In the following the connection of Combinatorial Hierarchy with genetic code is demonstrated.

2.1 Combinatorial Hierarchy As A Model For Abstraction Process

One could view the development of intelligence as a process, which takes place as sub-sequent transitions from hierarchy level to a higher hierarchy level, meta-level. First there are statements about concrete things, say numbers. Then come the statements about statements (say theorems about theorems of number theory): "If theorem A is true then theorem B is true". Then come the statements about statements about...

What is remarkable is that the so called Combinatorial Hierarchy [A3] (which emerged first in particle physics context) results from this kind of gradual abstraction process consider Z_2 valued functions. The value 1 might correspond to "true" and 0 to "not true".
i) Assume that lowest level corresponds to 2 statements.
ii) Go to the meta level and consider statements about statements (theorems about theorems in mathematics). Therefore one must consider Z_2 valued Boolean functions in Z_2 corresponding to statements of type "P is true" and "P is not true": there are altogether 4 of them. Drop the statement represented by $(0,0)$.
iii) Again one goes to meta level and considers statements about statements about statements that is functions from 3-element set to Z_2: 7 elements altogether, when one throws away the statement represented by sequence of zero bits.
iv) Continuing this process one clearly gets Combinatorial Hierarchy.

The somewhat mysterious feature is the dropping of one statement. If the construction corresponds to the construction assigning creation of the cognitive fermion pair with the splitting of a wormhole contact connecting two space-time sheets, then the requirement that the two space-time sheets form a connected structure drops exactly one configuration (no wormhole contacts connecting the two space-time sheets) from consideration. Second interpretation would be in terms of conscious bit. If conscious bit "1" is represented as a phase transition of some kind then the sequence of "0" s is not representable as a conscious bits so that only 2^k-1 code words are representable for k bit code and the only $k-1$ bits are fully representable which suggests that only 2^{k-1} bits represent information and remaining bits could play the same role as parity bits.

There is second construction which is purely set theoretical and gives a natural explanation for the dropping of one statement. Consider subsets of the set $(0,1)$. There are 3 of them if the physically non-realizable empty set is excluded. Consider next the subsets of the 3-element set: there are $2^3-1=7$ of these sets if empty set is excluded. By continuing the process one finds that the numbers of the Combinatorial Hierarchy result. This suggests that the physical non-realizability of $(0,0,...)$ Boolean statement is basic reason for dropping it from consideration.

The numbers of the Combinatorial Hierarchy have some properties, which suggest that they are very closely related to Genetic Code.

1. The numbers $p=3=M_2, 7=M_3, 127=M_7, M_{127}$ are Mersenne primes. It is possible that all the Mersenne numbers $M(n+1)=M_{M(n)}$ of the sequence are primes. This implies that the statements can be given the algebraic structure of Finite Field $G(p,1)$. Therefore the set of the Boolean statements has also interpretation as a simplified model for arithmetics and the arithmetic abilities of the intelligent system grow gradually in transitions from level to meta level. As a consequence the basic conjecture

$$M(n+1) \quad = \quad M_{M(n)}=2^{M(n)}-1 \text{ prime } n=1,2,... \ , \tag{2.1}$$

   roughly means that there is no upper bound for the arithmetic abilities of intelligent system!

2. The statements at the level p are of two types: "P is true" and "P is not true" and there are clearly $N(p)=(p+1)/2$ mutually consistent statements at level p. These numbers come as $2,4,64,2^{126},..$ for $p=3,7,127,...$ What is remarkable is that $N(7)=4$ is the number of different DNA:s and $N(127)$ is the number of different DNA sequences. This suggests that DNA: s represent physically consistent with a given atomic statement at level p.

n	1	2	3	4
$M(n)$	3	7	127	$2^{127}-1$
$N(DNA,n)=(M(n)+1)/2$	2	4	64	2^{126}
$N(amino,n)=(M(n)-1)/(M(n-1)-1)$	2	3	21	$(2^{126}-1)/63$

Table 1: The lowest Mersenne numbers of Combinatorial Hierarchy (known to be primes), the numbers of "DNA" k_{n-1}-plets and the numbers of "amino-acids" for these levels.

3. The finite field $G(p,1)$ has the cyclic group Z_{p-1} as the multiplicative group of nonzero elements. The dimensions of these groups read as $1,2,6=2\cdot 3,126=6\cdot 21,...$ for Combinatorial Hierarchy. The multiplicative group of the previous level can be regarded as a subgroup of the next level multiplicative group for the lowest members of the Combinatorial Hierarchy at least: $Z_{p_n-1}\subset Z_{p_{n+1}-1}$. The reason is that $p_{n+1}-1$ is divisible by p_n-1 for the lowest Mersenne primes. In fact divisibility holds true also for $M_{127}-1$ and M_7-1. Divisibility condition gives

$$2^{126}-1 \;=\; 63n \;. \tag{2.2}$$

The condition is satisfied!: $63=3^2\cdot 7$ divides $2^{126}-1$, whose prime factorization [A4] is given by

$$\begin{aligned} 2^{126}-1 &= 3^3\cdot 7^2\cdot 19\cdot 43\cdot 73\cdot 127\cdot X \;, \\ X &= 337\cdot 5419\cdot 92737\cdot 649657\cdot 77158673929 \;. \end{aligned} \tag{2.3}$$

Actually the divisibility follows quite generally from the following little theorem:
Theorem: $M(n)-1$ divides $M(n+1)-1$ always and $M(n)$ divides $M(n+1)-1$ if $M(n)$ is prime. The divisibility of $M(M(n))-1$ by prime $M(n)$ follows as a particular case $a=2, p=M(n)$ of Fermat's theorem stating that $a^{p-1}=1\ mod\ p$ holds true for any natural number a and any prime p. The divisibility of $M(M(n))-1$ by $M(n)-1$ is equivalent with the divisibility of $2^{M(n)-1}-1$ with $2^{M(n-1)-1}-1$. This property holds true for the lowest Mersenne numbers of the Combinatorial Hierarchy and can be proven to hold true generally by induction using the following lemma [A1]:
Lemma: The greatest common multiplier (x,y) for integers $x=2^a-1$ and $y=2^b-1$ satisfies $(2^a-1,2^b-1)=2^{(a,b)}-1$. The proof of the lemma is based on the observation that the polynomial x^n-1 ($x=2$ now) factorizes into a product of factors (x_i^n-1), where n_i is factor of n. For the polynomials x^a-1 and x^b-1 the largest common multiplier is therefore $x^{(a,b)}-1$.
For $a=M(n)-1$ and $b=M(M(n+1))-1$ lemma together with the induction assumption gives $(a,b)=2^{(M(n)-1,M(n-1)-1)}-1=2^{M(n-1)-1}-1$: the result means that $2^{M(n-1)-1}-1=M(n)-1$ divides $2^{M(n)-1}-1=M(n+1)-1$. The prime number property of Mersenne numbers is not needed in the proof.

As a consequence one has an infinite hierarchy of coset spaces $Z_{p_{n+1}-1}/Z_{p_n-1}=Z_{k_n}$, which are also cyclic groups. The first members of this hierarchy are $Z_{k_1)}=Z_2,Z_3,Z_{21},....$ Also the groups Z_{k_n} satisfy the condition $Z_{k_n}\subset Z_{k_{n+1}}$ for the lowest values of k with k_1 excluded and an interesting possibility is that k_n divides k_{n+1} quite generally for some number theoretic reason. What is remarkable is that $k_2=3$ is the number of DNA: s in DNA triplets and $k_3=21$ is the number of amino-acids plus stopping sign coded by DNA: s. This observation suggests that amino-acids correspond to the subset of the statements of $G(p_n,1)$ imbeddable as subgroup Z_{k_n} to Z_{p_n-1} for $p_n=127$. A little consideration shows that the map $x\to x^{k_{n-1}}$ gives a unique imbedding for the amino-acid type statements to Z_{p_n-1}.

These observations suggest a general model for Genetic Code. The 64 DNA sequences give a physical representation for the statements compatible with a given atomic statement at $p=127$ level of the Combinatorial Hierarchy. The choice of these mutually compatible statements as a subset of $G(127,1)$ is by no means unique and is determined by the evolution. The 21 amino-acids (stopping sign is regarded formally as "amino-acid") at $p=127$ level correspond to the unique

statements representable in the form $y = x^6$ in Z_{126} and form cyclic group Z_{21}. Genetic code can be regarded as the mapping $x \to x^6$ mapping all DNA type statements to amino acid type statements. The interpretation of the amino-acid type statements is as general axioms and DNA type statements are regarded as special cases of these axioms. Amino-acid sequences in turn are regarded as theorems derivable from axioms by constructing amino-acid sequences: the direction of the sequence is unique by the chirality of the amino-acid molecules.

The formation of theorems in formal system indeed corresponds to the formation of symbol sequences by some rules (now the rules are extremely simple, perhaps too simple!). The representability of DNA: s as triplets of 4 basic units has explanation: actually a more general the formula $N(DNA, n) = (p_n + 1)/2 = ((p_{n-1} + 1)/2)^{k_{n-1}} = N(DNA, n-1)^{N(amino,n-1)}$ holds true for all levels of the Combinatorial Hierarchy. In particular, for $p = 127$ one has $64 = 4^3$. Actually an infinite hierarchy of Genetic Codes suggests itself: for the n: th member of the Hierarchy there are k_n axioms and if k_n divides k_{n+1} the axioms of level n are imbeddable as a subgroup Z_{k_n} to the group $Z_{k_{n+1}}$ of axioms at level $n + 1$ and the counterpart of Gödel's Incompleteness Theorem holds true.

2.2 Interpretation Of Genetic Code

Finite Field Computer picture leads to the interpretation of genetic code as a map from the set of well defined truth values $X_{64} \subset G(127, 1)$ (DNA: s) to the set $Z_{126}/Z_6 = Z_{21} \subset G(127, 1)$ (amino-acids).

1. The interpretation of the previous observation is that amino-acids and "stopping sign" correspond to the elements x of the coset space $Y = Z_{126}/Z_6 = Z_{21}$ obtained by identifying two elements a and b of Z_{126} are identified if $a^6 = b^6$ holds true. The realization of Z_{21} as a subset of $G_{127,1}$ is obtained as the set of non-vanishing sixth powers of $G(127, 1)$ elements

$$z \in Z_{21} <=> z = x^6, \quad 0 \neq x \in G(127, 1) \ . \tag{2.4}$$

   Since a coset space is in question $d = 126/21 = 6$ elements of Z_{126} are mapped to a given element of Z_{21} in the map $x \to x^6$.

2. According to the proposed model of intelligent system DNA triplets correspond to a subset X_{64} of 64 well determined truth values of $p = 127$ logic but at this stage there is no first principle telling which subset corresponds to truth values. Let us however assume that X_{64} corresponds to subset of Z_{126} so that zero element is excluded from X_{64}.

3. With these identifications genetic code correspond to the mapping $x \to x^6$ from the set X_{64} of well defined truth values to Z_{21}

$$x \in X_{64} \quad \to \quad x^6 \in Z_{21} \ . \tag{2.5}$$

   The number of DNA triplets d coding same amino acid is just the number of elements of X_{64} mapped to same element of Y and this gives strong constraints for the identification of X_{64} as subset of $G(127, 1)$. Clearly, genetic code is to a high degree equivalent with the identification of 64-element DNA triplets as subset of Z_{126}. The identification of X_{64} as subset of Z_{126} cannot however be determined uniquely since the group Z_6 acts as the symmetry group of the code permuting the 6 elements of Z_{21} mapped to same element of Z_{126} and leading to a code with same degeneracies d.

An important prediction is that at most six DNA triplets can correspond to same amino acid. As**Table 2** shows, the condition is satisfied: there are three amino acids for which DNA degeneracy d is 6. This means that it is indeed possible to realize genetic code in the proposed manner.

One can consider first simple guesses for the identification of X_{64} as a subset of $G(127, 1)$. The identification as even elements $y = 2k$, $k = 0, 63$ is not possible since zero cannot belong to X_{64}:

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

Table 2: The number of amino acids N associated with a given degeneracy d telling the number of DNA triplets mapped to the amino acid in genetic code. The degeneracies are always smaller than 7 as predicted by the proposed explanation of the Genetic Code.

same applies to the identification as odd elements. The identification as the elements expressible as squares $y = x^2$ is excluded for the same reason. One could include the 64^{th} DNA by identifying it as an arbitrary element of Z_{126}. All these identifications yield almost completely symmetric genetic code: $d = 3$ for all 21 amino acids except one for with one has $d = 4$ so that something more complicated is needed.

2.3 Genetic Code As A Result Of Geometric Symmetry Breaking

One could try to understand the pattern of degeneracies as a symmetry breaking pattern, but not in terms of group representations as is done usually but in terms of group orbits. For single DNA multiplet associated with given amino-acid the natural symmetry group is Z_6 or some of its subgroups and DNA multiplet of given amino-acid can be regarded as a union of orbits for the subgroup in question. The subgroups of $Z_{126}/Z_6 = Z_3Z_7$ in turn can transform the DNA multiplets with same degeneracy to each other and amino-acids with same degeneracy can be regarded as a union of orbits of Z_7, Z_3 or Z_1. This symmetry pattern seems to work!

1. Consider first the group Z_6 acting inside DNA multiplets associated with given amino-acid. The definition of Z_{21} implies that the DNA: as associated with $d = 6$ amino-acids must be identified as six-element orbits of Z_6 symmetry. $d = 3$, $d = 2$, $d = 1$ amino acids correspond to the breaking of Z_6 to Z_3, Z_2 and to Z_1 respectively. $d = 4$ amino-acids are however problematic since Z_6 does not have 4-element subgroup. One can interpret $d = 4$ DNA: s either as a union of two Z_2 orbits or as a union of Z_3 and Z_1 orbits: Z_1 orbit corresponds most naturally to the mirror image of one Z_3 point, when the points of Z_{126} are represented as points of unit circle. This interpretation seems to be more appropriate.

2. Consider next the group Z_{21} acting on amino-acids. The 3 $d = 6$ amino-acids could be identified as an orbit of $Z_3 \subset Z_{21}$. If one regards the 5 $d = 4$ multiplets as unions of Z_3 and Z_1 orbits one can obtain altogether $2+5$ $Z_3 \subset Z_6$ orbits, which could be regarded as Z_7 orbit. What remains is $5+2 = 7$ Z_1 orbits, which could be regarded as Z_7 orbit. Therefore it seems that amino-acids can be ordered nicely into the orbits of Z_3 and Z_7. It turns however that exact symmetry is broken and some orbits are slightly deformed or even broken to pieces.

There are good reasons for symmetry breaking.

1. Besides symmetry also redundancy of the genetic code is desirable: this means that the number of the amino-acids with small degeneracy should be as small as possible. Z_6 (or some subgroup) symmetry and redundancy are competing factors since average degeneracy is always same and symmetry tends to increase the redundancy associated with some amino acids. Therefore redundancy requirement might be one underlying reason for symmetry breaking.

2. It turns out that also the competition between symmetries Z_3 and Z_7 in Z_{27} forces either Z_7 or Z_3 symmetry breaking already before the "actual" symmetry breaking.

3. The third reason for the symmetry breaking is the constraint that the numbers of DNA triplets and amino-acids are constrained to 64 and 21 respectively. For, instance 63 DNA: s allows representation as a union 9 Z_6 orbits having $Z_3 \subset Z_{21}$ as symmetry group but for 64 DNA: s there is necessarily one Z_1 orbit present.

It would be nice if one could understand the genetic code as a small perturbation of some code with both high symmetry and high redundancy. One could even assume that the number of amino-acids before the symmetry breaking is smaller than 21 so that symmetry breaking is necessary to obtain 21 amino-acids. A geometric picture of the situation is obtained by regarding the points of Z_{126} as points $\Phi = n2\pi/126$ of a unit circle endowed with standard metric so that it becomes possible to define what "small" symmetry breaking means.

2.4 Symmetry Breaking Scenarios

The genetic code can be reproduced as the following symmetry breaking pattern.

1. Unbroken symmetry corresponds to the following situation. There are $6 = 3 + 3$ amino-acids with maximal degeneracy $d = 6$ and 7 amino-acids with degeneracy $d = 4$: altogether $13 < 21$ so that symmetry breaking is necessary. $d = 6$ multiplets correspond to 6 Z_6 orbits $\Phi = n2\pi/6 + k\Delta(i)$, with $\Delta(i) = k_i 2\pi/126 < 2\pi/6$. Single $d = 4$ multiplet corresponds to Z_3 orbit plus single point, which is diametrically opposite to one of the points at Z_3 orbit. More explicitly: basic $(3, 1)$ multiplet corresponds to the Z_3 orbit $\Phi_3(k) = 6\Delta + k2\pi/3$, $k = 0, 1, 2$ plus the point $\Phi_1 = 6\Delta + \pi$. By acting on this orbit with rotations $\Delta_1(i) = k_i 2\pi/126$, $i = 0, ..., 6$, $\Delta_1(i) < 2\pi/6$ one obtains 7 $d = 4$ multiplets. Obviously one must have $\Delta(i) \neq \Delta_1(j)$ for each i, j pair in order to avoid overlapping. The actual imbedding of Z_6 multiplets is not relevant for the degeneracies of the genetic code and will be discussed later.

2. The first symmetry breaking is $Z_6 \to Z_2$ and occurs for 3 $d = 6$ amino-acids and leads from 3 Z_6 orbits to 9 Z_2 orbits so that 9 $d = 2$ amino-acids result. Breaking can be understood geometrically as follows. Single $d = 6$ multiplet is obtained by Z_2 action (reflection) from Z_3 orbit, say $\Phi(k) = k2\pi/3$. What happens is that two points on Z_3 orbit that for Z^3 orbit the points $\Phi(k)$ with $k = 1$ and 4 are rotated slightly with different rotation angles

$$\begin{aligned} \Phi(1) &\to \Phi(1) + \delta_1 \ , \\ \Phi(4) \to \Phi(4) + \delta_1 \ , & \\ \Phi(5) &\to \Phi(5) - \delta_2 \ , \\ \Phi(6) &\to \Phi(6) - \delta_2 \ , \\ \delta_i &= \frac{k_i 2\pi}{126} \ , \\ 1 < k_i &< 6 \ . \end{aligned} \tag{2.6}$$

   Here δ_i must be chosen so that the deformed points do not coincide with already "occupied" points. Minimal symmetry breaking is obtained with $k_1 = 1 = k_2$ but would lead to overlapping. One must also have $k_1 \neq k_2$ ($k_1 = k_2$ would imply additional reflection symmetry). The symmetry breaking leads to $9 + 7 = 16$ amino acids with degeneracies $d \geq 2$.

3. The second symmetry breaking leads from 7 $d = 4$ multiplets to 5 $d = 4$ multiplets, 2 $d = 3$ multiplets and 2 $d = 1$ multiplets. What happens is that 2 basic $d = 4$ multiplets consisting of Z_3 orbit and mirror image of one of Z_3 points is deformed $d = 3$ plus $d = 1$ multiplet: this is achieved if mirror image point is slighly shifted. The deformation is obtained by performing for the basic multiplet $(\Phi_3(k) = k2\pi/3, \Phi_1 = \pi)$ the deformation

$$\Phi_1 = \pi \quad \to \quad \pi \pm \delta_3 \ . \tag{2.7}$$

   $\delta_3 = k_3 2\pi/126$, $1 < k_3 < 6$ followed by appropriate rotation carrying broken multiplet to its own position. $k_3 = 1$ is not allowed since it would lead to overlapping and $k_3 = 2$ leads to the smallest possible symmetry breaking.

One can try to find unique imbedding of X_{64} (and also try to understand the uniqueness of the genetic code) by requiring that amino-acids form orbits of Z_{21} or its subgroups. Z_3 and Z_7

d	N(d)	$\Phi/\Delta = n \; mod \; 126$, $\Delta = 2\pi/126$
6	3	$n = A(i) + B(k)$, $A(i) = 21i$, $B(k) = 7k$, $i \in I$, $k \in K$
		$I = \{0, 1, .., 5\}$, $K = \{0, 1, 2\}$
2	9	$n = (A(i) + B(k) + 1 + \delta_1(i))$, $i \in I$, $k \in K$
		$\delta_1(0) = \delta_1(3) = 0$, $\delta_1(1) = \delta_1(4) = -3$, $\delta_1(2) = \delta_1(5) = 4$
4	5	$n = i + 42k$, $n = i + 63$, $i = 2, 10, 11, 17, 18$, $k \in K$
3	2	$n = i + 42k$, $i = 3, 4$, $k \in K$
1	2	$n = i + 63 + \delta_2(i)$, $i = 3, 4$, $\delta_2(3) = -1$, $\delta_2(4) = 1$

Table 3: Explicit form for the Z_3 symmetric imbedding of X_{64} consistent with Genetic Code. The index k appearing in the formulas labels points on Z_3 orbit. The imbedding is illustrated in **Fig. 1**.

d	N(d)	$\Phi/\Delta = n \; mod \; 126$, $\Delta = 2\pi/126$
6	3	$n = A(i) + B(k)$, $A(i) = 21i$, $B(k) = 7k + \delta_1(k)$, $i \in I$, $k \in K$
		$I = \{0, .., 5\}$, $K = \{0, 1, 2\}$, $\delta_1(1) = 5$, $\delta_1(0) = delta_1(2) = 0$
2	9	$n = A(i) + \delta_2(i) + 1 + B(k)$, $i \in I$, $k \in K$
		$\delta_2(1) = \delta_2(4) = -2$, $\delta_2(2) = \delta - 2(5) = 2$,
4	5	$n = 3 + i + 42k$, $n = i + 63$, $i = 2, ..., 6$, $k \in K$
3	2	$n = 3 + i + 42k$, $i = 1, 7$, $k \in K$
1	2	$n = i + 63 + \delta_3(i)$, $i = 1 \; or \; 7$, $\delta_3(1) = -1$, $\delta_3(7) = 1$

Table 4: Explicit form for the Z_7 type imbedding of X_{64} consistent with Genetic Code. The imbedding is illustrated in figure **??**.

are competing symmetries and not consistent with each other: the reason is that Z_3 orbits and Z_7 orbits in same basic angular range of length $2\pi/6$ necessarily overlap since between two points of Z_3 orbit there are always just 6 points so that Z_7 orbit cannot be put between two points on Z_3 orbit. Therefore one must choose between either enhanced Z_3 or Z_7 symmetry for the imbedding of X_{64}.

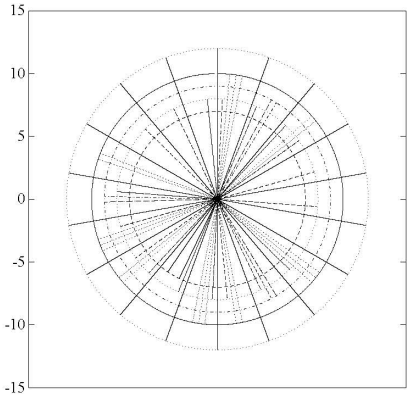
Consider first Z_3 symmetric situation.

1. $6 = 3+3$ Z_6 orbits decompose naturally into two orbits of Z_3. The first Z_3 orbit corresponds to $\Phi(k, m) = 6m \cdot 2\pi/126 + k2\pi/6$, $m = 0, 1, 2$ and second Z_3 orbit is obtained from this orbit by a rotation $\Delta\Phi = 2\pi/126$. When Z_6 orbits break down to 9 Z_2 orbits the Z_2 orbits can form 3 separate orbits of Z_3.

2. The graphical experimentation with various possibilities shows that one must split Z_7 orbit to three pieces $7 = 1 + 2 + 2 + 2$ so that the 2: s form Z_3 orbit. Therefore complete breaking of Z_7 symmetry makes possible additional Z_3 symmetry. This means that one $d = 4$ orbits arrange into 2 Z_3 orbits and one Z_1 orbit before symmetry breaking. After the symmetry breaking the 2 $d = 3$ multiplets and $d = 3$: s belonging to 5 $d = 4 : s$ form still Z_3 orbits but two $d = 1$: s are thrown out of corresponding Z_3 orbits.

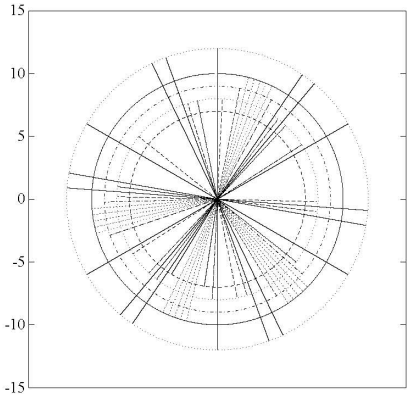
3. An example of an imbedding satisfying these constraints is given by the following formulas

Consider next Z_7 type scenario. 7 $d = 4$: s can be put into single Z_7 orbit. $d = 6$ representations cannot however form neither Z_7 full orbits nor full Z_3 orbits in this case. After the symmetry breaking the 2 $d = 3$ multiplets and $d = 3$: s belonging to 5 $d = 4 : s$ form still Z_7 orbits but two $d = 1$: s are thrown out of corresponding Z_7 orbits. There exists no full Z_7 orbit consisting of amino-acids with same d after symmetry breaking so that in this sense Z_7 sceneario possesses much less symmetry than Z_3 scenario. An explicit example is given by the following formulas



1

Figure 1: The imbedding $X_{64} \subset Z_{126}$ reproducing Genetic Code and possessing Z_3 type symmetry. The lengths of radial lines are $6+d$, where $d = 1, 2, 3, 4, 6$ is the number of DNA: s associated with amino-acid. The angular distance between points on Z_3 (Z_7) orbits is to 20 (2.85) degrees.



1

Figure 2: Z_7 type imbedding $X_{64} \subset Z_{126}$ reproducing Genetic Code. Symmetry breaking is much larger for this imbedding although visually the imbedding looks perhaps more symmetric than Z_3 type imbedding.

$N(d=1)$	$N(d=2)$	$N(d=3)$	$N(d=4)$	$N(d=6)$
0	12	0	7	2
2	**9**	**2**	**5**	**3**
4	6	4	3	4
6	3	6	3	5

Table 5: The 4 possible genetic codes assuming unbroken initial Z_3 symmetry with numbers of amimo acids with same degeneracy.

2.5 In What Sense The Physical Genetic Code Is Unique?

The proposed symmetry breaking scenario is certainly not the only one. A constraint on symmetry breaking patterns comes from the requirement that all amino acids are coded. In terms of orbit multiplicities $g(k)$ (number of amino acids with same number of DNA: s) one has two conditions

$$\begin{aligned} \sum_{k=1,2,3,4,6} g(k) &= 21 \ , \\ \sum_{k=1,2,3,4,6} g(k)k = 64 \ . \end{aligned} \tag{2.8}$$

$k = 4$ case corresponds to two Z_3 and Z_1 orbits associated with single amino acid. This gives

$$\begin{aligned} g(1) &= -22 + g(3) + 2g(4) + 4g(6) \ , \\ g(2) &= 43 - 2g(3) - 3g(4) - 5g(6) \ . \end{aligned} \tag{2.9}$$

This means that all possible genetic codes are labeled by the three integers $g(3), g(4)$ and $g(6)$. The conditions imply

$$\begin{aligned} g(3) + 2g(4) + 4g(6) &\geq 22 \ , \\ 2g(3) + 3g(4) + 5g(6) &\leq 43 \ . \end{aligned} \tag{2.10}$$

These conditions restrict the possible symmetry breaking scenarios. In particular, $g(6) \leq 8$ must hold true so that maximal symmetry corresponds to the codes with $(g(3), g(4), g(6))$ equal to $(1, 0, 8)$ or $(0, 1, 8)$.

In the proposed symmetry breaking scenario the number of DNA: s is automatically fixed to 64 and the only requirement is that deformation of X_{64} is such that the number of amino-acids is not smaller than 21. If one assumes that the only symmetry breakings are of form $6 \to 3 \cdot 2$ and $4 \to 3 + 1$ and denotes the numbers of broken $d = 6$ and $d = 4$ multiplets with k_6 and k_4 respectively the condition for 21 amino acids reads as $(6 - k_6) + 3k_6 + (7 - k_4) + 2k_4 = 21$, which gives $k_6 = 4 - k_4/2$, $k_4 = 0, 2, 4, 6$. k_4 gives the number of $d = 1$ amino-acids and $k_4/2 + 2$ gives the number of the unbroken 6 orbits: these numbers clearly measure redundancy and symmetry. The numbers of amino-acids with various degeneracies are given by

$$\begin{aligned} N(d=1) = N(3) &= k_4 \ , \\ N(2) &= 3(4 - k_4/2) \ , \\ N(4) &= 7 - k_4 \ , \\ N(6) &= 2 + k_4 \ . \end{aligned} \tag{2.11}$$

Table 5 summarizes the 4 genetic codes possible under these assumptions.

The physically realized genetic code (second row) is clearly a compromise between symmetry and redundancy. By the previous construction the physically realized code is characterized by additional symmetry: namely the group $Z_3 \subset Z_{21}$ transforming both 3 $d = 6$ amino-acids and

3 triplets of $d = 2$ amino acids to each other. For all other Z_3 type alternatives this symmetry is broken. Therefore the physical genetic code corresponds in a well defined sense to minimal symmetry breaking.

2.6 Hierarchy Of Genetic Codes?

Genetic Code generalizes to an entire hierarchy of genetic codes at formal level, at least.

1. The spaces $X_4 \subset X_{64} \subset X_{(M_{127}+1)/2} = X_{2^{126}} \subset ....$ an be regarded as a hierarchy of "DNA triplets".

2. The preceding results suggest that the multiplicative groups associated with the Combinatorial Hierarchy form also a hierarchy: $Z_2 \subset Z_6 \subset Z_{126} \subset Z_{M_{127}-1} \subset .....$ This is true if the dimensions divide each other: $2|6|126|M_{127}-1|.....$ 126 indeed divides $M_{127}-1 = 2(2^{126}-1)$ and the number of "amino-acids" at the third level is $(2^{126}-1)/63$! The result holds generally from Fermat's theorem ($2^{p-1} = 1 \ mod \ p$) if the Mersenne numbers of the Combinatorial Hierarchy are primes.

3. If Combinatorial Hierarchy consists of primes the coset spaces $Z_6/Z_2 = Z_3 \subset Z_{126}/Z_6 = Z_{21} \subset Z_{M_{127}-1}/Z_{126} \subset ....$ exist and form a hierarchy of "amino-acids".
   The redundancy of the genetic code could be interpreted as resulting from a hierarchy of discrete gauge symmetries: Z_{M_n-1} acts as gauge symmetry in the space $Z_{M_{M_n}-1}$.

There is still further apparent numerical co-incidence. For the first 3 levels the numbers of "amino-acids" are

$$\begin{aligned} a_0 &= 2 \ , \\ a_1 &= 3 \end{aligned} \tag{2.12}$$

$$\begin{aligned} per, & \\ a_2 &= 21 \ , \\ a_3 &= ((M_{127}-1)/2)/((M_7+1)/2)) = (2^{126}-1)/63 \ . \end{aligned} \tag{2.13}$$

The numbers of "DNA" -n-plets (well defined truth values) are

$$\begin{aligned} d_0 &= 2 \ , \\ d_1 &= 4 = 2^2 = d_0^{a_0} \ , \\ d_2 &= 64 = 4^3 = d_1^{a_1} \ , \\ d_3 &= 2^{126} = 64^{21} = d_2^{a_2} \ . \end{aligned} \tag{2.14}$$

The formulas imply that one can construct the physical DNA-triplets at level 2 as $d_2 = d_1^{a_1} = 4^3$ triplets of DNA molecules. At level 3 one can construct $d_3 = d_2^{a_2} = 63^{21} = 2^{126}$ "DNA" 21-plets of 64 different DNA-triplets. More generally, DNA-plets at given level n correspond to a_{n-1}-plets of DNA formed from the DNA-plets of the previous level since the identity

$$d_n = \frac{(M_n+1)}{2} = d_{n-1}^{a_{n-1}} = 2^{a_{n-1}a_{n-2}a_{n-3}...a_0} \ . \tag{2.15}$$

holds true in general since the groups Z_{M_n-1} forms a hierarchy of nested subgroups.

The result means that the concrete physical representation of truth values of M_n logic as "DNA" sequences is an internal property of the Combinatorial Hierarchy. The immediate prediction is that sequences of 21 DNA-triplets should code for $a_3 = (2^{126}-1)/63$ generalized amino-acids consisting of sequences of 21 amino-acids acids: these units could be regarded as some kind of "pre-genes". The number of amino-acid sequences consisting of 21 amino-acids (of order 2^{91} is much smaller

than the number of generalized amino-acids (of order 2^{120}), which means that the ordinary amino acids cannot provide an optimal realization of DNA sequences: many generalized amino-acids have no DNA sequence as their representative. DNA sequences are known to contain passive sections, introns, which do not code DNA. An interesting question is whether these sections could represent M_{127} level sequences: if so the number of DNA triplets in these sections should be a multiple of 21.

If the lowest level of the Hierarchy is present DNA: s should be representable as doublets formed from two different $p = 3$ "pre-DNA": s and these should code for 3 different "pre-amino-acids". This kind of structure is not realized in Nature so that the $p = 7$ level of DNA: s is in this sense irreducible.

2.7 The Structure Of The Negation Map

The negation map mapping statement to its negation is highly non-unique unless it possesses symmetries and it is interesting to find what one can conclude about the structure of this map by symmetry arguments.

1. For each amino-acid type general statement there are six special cases of the statement. The fact that only 64 statements are actually consistent with a given atomic statement means that some special cases associated with different amino-acids correspond to statement and its negation whereas two points at given Z_6 orbit cannot correspond to a statement and its negation.

2. 62 negations of 64 DNA type statements belong to Z_{126}. One lacking negation corresponds to the zero element of $G(127, 1)$ and the second one corresponds to the excluded statement in the construction of the Combinatorial Hierarchy. One can associate to each amino-acid definite number of negations as the number of elements in the complement of DNA type statements on the Z_6 orbit defined by amino-acid (at Z_6 orbit all statements must be consistent with a fixed atomic statement). These numbers are $d_1 = 0$ for 3 $d = 6$ amino-acids, $d_1 = 2$ for 5 $d = 3 + 1$ amino-acids, $d_1 = 4$ for 9 $d = 2$ amino-acids and $d_1 = 5$ for the 2 $d = 1$ amino-acids so that negations "code" only 18 amino-acids so that duality symmetry between statements and their negations is not possible.

3. Negation must map statements at given orbit to a different orbit. It is however natural to require that points on same orbit, say A, are mapped on same orbit, say B, if possible and that the angles and ordering of points belonging to same orbit are preserved in negation map. More generally, the ordering of points on circle should be preserved.

4. For 2 $d = 3$ amino-acids, call them a and b, the Z^3 orbit a (b) is mapped to the complement of Z^3 orbit b (a). The map is unique apart from Z^3 rotation.

5. For 9 $d = 2$ amino-acids Z^2 orbits (pairs of diametrically opposite points) form 3 Z^3 orbits and Z^3 rotation plus a rotation inside Z^6 orbit gives good candidate for negation map. As a consequence 9 Z_2 orbits in the complement remains "free" and form 3 Z_3 orbits, call them G_i.

6. It is natural map $d = 6$ orbits to these free 9 Z_2 orbits in Z_3 symmetric manner, which means that the content of single Z_6 orbit i is mapped into 3 $d = 2$ orbits in G_i. Diametrically opposite point pair on Z_6 orbit is mapped to a similar point pair in G_i. The map is unique if one requires that Z_6 element on the initial orbit goes to same Z_6 element on the final orbit and this in turn makes the map of 9 Z_2 orbits unique apart from reflection.

7. The 5 $d = 3 + 1$ amino-acids form unions of diametrically opposite point pair A_i plus point pair B_i, $i = 1, ..., 5$ separated by angle of 120 degrees. Point pairs B_i can be mapped to the mirror image $\bar{B}_j$ of B_j, $j \neq i$: one can require that cyclic ordering of pairs B_i is preserved in map to remove part of the non-uniqueness. $2 + 2$ pairs A_i can be mapped to the $2 + 2$ pairs C_i in the complements of 2 $d = 1$ orbits containing 5 points each (2 diametrically opposite pairs C_i plus single point). The remaining pair, say A_{i_0}, can be mapped to zero element of $G(127, 1)$ and to the "dropped out" statement.

8. Denoting the mirror images of 2 $d = 1$ statements a and b by $\bar{a}$ and $\bar{b}$ the negation map for them reads as $(a, b) \to (\bar{b}, \bar{a})$.

2.8 Combinatorial Hierarchy As A Hierarchy Of Formal Systems

Usually [A2] formal system is understood as a system of symbols, axioms interpreted as allowed basic strings of symbols and rules for constructing new sequences from the symbols. In [A2] the exciting analogies between the symbol sequences of formal systems and DNA and amino acids sequences and Genetic Code were emphasized and it indeed seems that more than analogy is in question. The model for the Genetic Code suggests an interpretation of Combinatorial Hierarchy as a hierarchy of formal systems with DNA type statements identifiable as a maximal set of 64 statements consistent with an atomic statement (single bit fixed) and amino-acid type statements as basic axiom type truths. Genetic code results from the necessary but non-unique selection of these statements consistent with a fixed atomic statement and the selection of these statements could be a result of fight for survival at the level of amino-acids. The rule for forming statements in this system is simple: just form DNA sequences with building blocks consisting of a_{n-1}(number of "amino-acids" at level $n-1$) DNA-sequences of previous level. Theorems are obtained by coding these statements to amino-acid sequences.

Consider first DNA type statements.

1. It was already suggested that the elements of Finite Fields in the Combinatorial Hierarchy correspond to a hierarchy of statements about statements about.... and therefore a sequence of formal systems formed from basic formal system by successive construction of meta level statements.

2. Finite Field provides a language to construct all possible statements. As already found, the $(p+1)/2$ (p is Mersenne prime in the Combinatorial Hierarchy) statements correspond to maximal set of statements consistent with a fixed atomic statement of this formal system. This is indeed a possible interpretation. Combinatorial Hierarchy was constructed by starting from a set containing just two statements 1 and 0. In the first $p = 3$ level one forms 4 statements about these two statements and drops the one representable as $(0, 0)$. There are however 2 contradictory statements of type P and *not* P in this set so that only $2 = (p+1)/2$ statements are consistent with a given atomic statement in the real world. At the level p one has $p = M_n$ statements about lower level statements and by construction $(p-1)/2$ statements P have also their negation in the set of all possible statements so that $(p+1)/2$ consistent statements with a fixed atomic statement are possible. Thus the conclusion is that 64 DNA triplets represent the maximum number of mutually consistent statements at level $p = M_7 = 127$ of a formal system possibly having a hierarchy of meta levels.

Consider next the interpretation for the set $Z_{k_n} = Z_{p_n-1}/Z_{p_{n-1}-1}$, $k_n = (p_n - 1)/(p_{n-1} + 1)$ of amino-acid type statements.

1. In any formal system there are two kind of meta statements that is statements $P(Q_1, ..Q_n)$ about statements. The first class corresponds to theorems $P(Q_1, ...Q_n)$ without any reference to the particular value of statements Q_i: a simple example is general theorem stating the conditions under which an orthogonal triangle with two integer sides is Pythagorean triangle: $m^2+n^2 = l^2$ with m, n, l integers. Second class corresponds to the theorems with $(Q_1, ..., Q_n)$ possessing definite values: a simple example is previous theorem for triangle having sides $2, 1$ and 3: $2^2 + 1^2 = 3^2$. The set of "amino-acid" type statements does not depend on the choice of the choice fo the $(p+1)/2$ statements consistent with a given atomic statement and this suggests the interpretation "amino-acid" type statements as general axiom like truths without any reference to the values of the argument of the statement.

   What is special in the proposed hierarchy of formal systems that the substitution operation corresponds to the multiplication of Z_k element with the element $Z_{p_{n-1}-1}$. The introduction of mutual consistency by selecting $(p+1)/2$ special statements implies that the number of DNA: s per amino-acid corresponds to the number of special cases associated with a given "axiom" depends on the "axiom". There is still upper bound for special cases given by $p_{n-1} - 1$.

2. In a good formal system the choice of DNA type statements should be such that there is at least one statement per each truth. The most important truths (as far as survival is considered) should correspond to full $Z_{p_{n-1}-1}$ orbits.

Gödel's Incompleteness Theorem was one of the basic themes in the book of Hofstadter [A2]. Theorem states that in any sufficiently complicated formal system (, that is, practically interesting) there are truths, which are not provable. The Incompleteness Theorem seems to have its analog in bio-systems. As already observed at the level M_{127} DNA sequences consisting of 21 triplets correspond to true statements, which can be regarded as special cases of general truths, whose number is $(2^{126-1})/63 \sim 2^{120}$. Amino-acid sequences consisting of 21 amino-acids give a natural realization for a subset of these truths and the genetic code map is induced from the Genetic Code at the basic level $p = 127$. The number of the truths given by these sequences is of order $20^{21} \sim 2^{91}$ (taking into account the reduction caused by the stopping sign) and much smaller than all possible truths. The interpretation is that the number of possible theorems obtained by forming amino-acid sequences is much smaller than the number of truths. One could always add "axioms" by realizing the remaining truths in some new manner but since the number of levels in the Combinatorial Hierarchy is infinite (assuming that the Mersenne numbers in question are primes) there are always unprovable truths in the system. One can obviously classify the formal systems according to which level is the basic irreducible level inducing genetic code at higher levels.

$p = 7$ ($n = 2$) and $p = 3$ ($n = 1$) level as the defining, irreducible level are also in principle possible.

1. At level $p = 7$ 4 doublets of level $p = 3$ "DNA" code 2 "pre-amino-acids" plus stopping sign and at the next level this coding induces $2^3 = 8 < 21$ different "amino-acids" since stopping sign does not appear in theorems. If stopping sign would correspond to actual amino-acid there would be $3^3 = 27 > 21$ theorems so that the number of theorems would be larger than the number of truths!

2. At level $p = 3$ 2 "DNA": s of code one amino-acid plus stopping sign so that genetic code is one-to-one.
   An interesting possibility is that $p = 3$ and $p = 7$ levels might have been defining, irreducible levels for bio-systems at some early stage of evolution. These formal systems can be regarded as a subsystem of the full DNA-amino-acid system. RNA-triplets of form UXY, $X, Y \in \{A, C\}$ give indeed realization for $p = 7$ system: UAA codes stopping sign, UAC codes cys and UCA and UCC code ser (using the standard short hand notation for RNA: s and amino-acids [A2]). The sequences formed from these DNA: s and corresponding amino-acids indeed realize $p = 7$ formal system as subsystem of $p = 127$ system. $p = 3$ formal system can be realized as UAA coding stopping sign and UAC coding cys. An interesting possibility is that this DNA-amino-acid subsystem has formed first in the biochemical evolution. For both $p = 3$ and $p = 7$ degenerate genetic codes one has $G + C/A + U = 1/2$. $G + C/A + U$ content varies in the range $(.7, 1.5)$ for insects and in the range $(1.17, 1.56)$ in the case of fish and for younger evolutionary forms A+U content is known to increase [I19].

3. The transition to "$p = M_{127}$ life" would require the addition of a rather large number of new "amino-acids" to the set of all possible amino-acid sequences consisting of 21 amino-acids whereas DNA triplets could be simply replaced with sequences of 21 DNA triplets. In the transition to $p = 127$ life tRNA triplets binding single amino-acid would be replaced by sequences of 21 tRNA triplets binding besides amino-acid sequences suitably modified amino-acid sequences in order to achieve maximal number of "amino-acids" at level $p = 127$. Also the modification of the translation system (in ribosomes) is required so that the "reading head" recognizes a sequence of 21 mRNA-triplets instead of single mRNA triplet. An interesting question is whether biochemistry might allow this kind of extension.

2.9 Summary

The simple model of abstraction has rather interesting connections with genetic code.

1. Combinatorial Hierarchy results as hierarchy of abstraction levels for statements or thoughts. Lowest level A_2 corresponds to the two possible values of Boolean variable and thoughts of

the level A_{n+1} correspond to statements about statements of level n that is Boolean maps $A_n \to A_2$. If the statement corresponding to sequence of zero bits is excluded the dimensions form a series of Mersenne numbers $3, 7, 127, M_{127}, ...$ Combinatorial Hierarchy is obtained also by considering the set of subsets with empty set excluded. The hypothesis that there is no upper bound for intelligence is equivalent with the hypothesis that these numbers are primes and that the numbers $p_n - 1$ in the sequence have the property that $p_n - 1$ divides $p_{n+1} - 1$. This implies that one obtains a hierarchy of Finite Fields $G(p_n, 1)$ and their multiplicative groups Z_{p_n-1} as well as coset groups $Z_{p_n-1}/Z_{p_{n-1}-1} = Z_{k_n}$, $k_n = (p_n - 1)/(p_{n-1} - 1)$.

2. There are $(p_n + 1)/2$ statements consistent with a given atomic statement at level n and the numbers come as $2, 4, 64, ...$ These statements are referred to as "DNA" type statements for obvious reasons. The dimensions k_n comes as $2, 3, 21, ...$ The hypothesis is that amino-acid type statements correspond to the statements of $Z_{k_n} \subset Z_{p_n-1}$ and can be regarded as general theorems whereas DNA type statements correspond to special cases of these theorems and are mapped to general theorems the mapping $x \to x^{p_{n-1}-1}$ at level n. At level n the genetic code corresponds to the non-unique choice of the $(p_n + 1)/2$ DNA type statements consistent with a given atomic statement in Z_{p_n-1}.

3. Biologically Genetic Code is determined by the competition between amino-acids: each amino-acid tries to maximize the number of DNA: s coding it (amino-acids are like politicians who are representatives of one truth and DNA: s are in the role of voters). The tendency favors $Z_{p_{n-1}-1}$ (Z_6) symmetry. The physically realized code can be understood as resulting from the symmetry breaking caused by the competition between the groups $Z_{p_{n-1}-1}$ (Z_6) acting on DNA: s associated with single amino-acid and factor groups of Z_{k_n} $(Z_{21} = Z_3 Z_7)$ transforming amino-acids to each other. Instead of finite dimensional representations of Lie groups the orbits of the cyclic groups Z_n are basic objects in the symmetry breaking mechanism.

4. At level n basic objects are k_{n-1}-plets of DNA: s of level $n - 1$ and sequences of same DNA units can in principle appear at all levels of the hierarchy. At the next $k = M_{127}$ level "DNA": s could therefore be regarded as sequences of 21 DNA triplets.

5. A hierarchy of increasingly complicated formal systems is predicted if one accepts the hypothesis. For the formal system of order n n: th level of the Combinatorial Hierarchy is the defining level in the sense that the number of "amino-acids" is maximal and equal k_n. Formal systems of order $n_1 < n$ are imbeddable into the formal system of order n. The formation of amino-acid sequences corresponds to the formation of theorems. For a formal system of order n amino-acid sequences realize only a small subset of all possible k_{n_1} truths at higher levels $n_1 > n$ of the Combinatorial Hierarchy in accordance with Gödel's theorem. One can also classify all possible bio-systems according to the value of n for the corresponding formal system. The Earthly life corresponds to $n = 3$ formal system and "life of order 4" would require the addition of rather large number of new "amino-acids" to the set of all possible amino-acid sequences consisting of 21 amino-acids whereas DNA triplets could be simply replaced with sequences of 21 DNA triplets. The realization of "$n = 4$ life" requires that tRNA triplets binding single amino-acid are replaced by sequences of 21 tRNA triplets binding besides amino-acid sequences suitably modified amino-acid sequences in order to achieve maximal number of "amino-acids" at level $n = 4$. Also the modification of the reading system (in ribosomes) is required so that the "reading head" recognizes a sequence of 21 mRNA triplets instead of single mRNA triplet.

6. An open problem relates to the precise role of DNA and proteins. The model of Boolean thoughts represented in terms of the cognitive fermion pairs leads to the correspondence between fermions and mind like space-time sheets and gives Combinatorial Hierarchy a special status. What comes in mind is that DNA provides a hardware representations of thoughts analogous to a computer memory. DNA molecules would be conscious selves representing 3 basic symbols in the mind of higher level self formed by DNA triplet. DNA sequences would be selves, experiencing DNA triplets as their sub-selves. Individual DNA molecules would represent sub-sub-selves so that DNA sequence would experience only the average of the experiences of individual DNA molecules.

7. It deserves to be noticed that I Ching claims that there are 64 fundamental mental states: could it be that these mental states correspond to all possible DNA triplet selves? It this interpretation is correct then Buddhist meditators would have achieved bio-feedback at DNA level! Genetic Code itself could be interpreted as a mapping of DNA selves to protein selves: this could be perhaps regarded as kind of mimicry or conscious abstraction process. Protein selves would represent theorem like abstractions of conscious thoughts represented by DNA selves.

8. It is known that cell numbers of different cell types in nervous, muscular, adipose, gonadic and homopoietic organs concentrate themselves around powers of two – 2^n, where n in the range $20-40$ [I22]: this can be understood if they result in n regularly occurring cell divisions. It might however be that the explanation of the regularity involves something much deeper. For instance, the cell types represent various n-bit sequences.

3 Combinatorial Hierarchy: two decades later

Combinatorial Hierarchy (CH) [A3, A5] is a hierarchy consisting of Mersenne integers $M(n) = M_{M(n-1)} = 2^{M(n-1)} - 1$ and starting from $M_1 = 2$. The first members of the hierarchy are given by $2, 3, 7, 127, M_{127} = 2^{127} - 1$ and are primes. The conjecture of Catalan is that the hierarchy continues to some finite prime. It was proposed by Peter Noyes and Ted Bastin that the first levels of hierarchy up to M_{127} are important physically and correspond to various interactions (see `http://tinyurl.com/hszo9wb`). I have proposed the levels of CH define a hierarchy of codes containing genetic code corresponding to M_7 and also memetic code assignable to M_{127} [K5].

Pierre Noyes and Ted Bastin proposed also an argument why CH contains only the levels mentioned above. This has not been part of TGD view about CH: instead of this argument I have considered the possibility that CH does not extend beyond M_{127}. With the inspiration coming from email discussion I tried to understand the argument stating that CH contains M_{127} as the highest level and ended up with a possible interpretation of the condition. Zero energy ontology (ZEO) and the representation of quantum Boolean statements $A \to B$ as fermionic parts of positive and negative energy parts of zero energy states is essential. This led to several interesting new results.

1. To my best understanding the original argument of Noyes does not allow M_{127} level whereas prime property allows. States at M_{127} level cannot be mapped to zero energy states at M_7 level. Allowing a wild association with Gödel's theorem, one could say that that there is hube number of truths at M_{127} level not realizable as theorems at M_7 level.

   A possible interpretation is that M_{127} level corresponds to next level in the abstraction hierarchy defined by CH and to the transition from imbedding space level to the level of "world of classical worlds" (WCW) in TGD. The possible non-existence of higher levels (perhaps implied if $M_{M_{127}}$ is not prime) could be perhaps interpreted by saying that there is no "world of WCWs"!

2. Rather remarkably, for M_7, which corresponds to genetic code [K5], the inequality serving as consistency condition is saturated. One can say that any set of 64 statements consistent with a given atomic statement (1 bit fixed) at M_7 level can be represented in terms of 64 Boolean maps at M_3 level representable in terms of zero energy states. One obtains an explicit identification for the Boolean algebras involved in terms of spin and isospin states of fermions in TGD framework at level M_7 so that genetic code seems to be realized at the fundamental elementary particle level thanks to the dimension $D = 8$ of imbedding space. Even more, the level M_{127} corresponding to memetic code emerges in the second quantization of fermions at M_7 level. Here color triplet property of quarks and color singletness of leptons and the identification of elementary particles as pairs of wormhole contacts are in essential role.

The conclusion would be that in TGD Universe genetic code and its memetic counterpart are realized at the level of fundamental particles. Already earlier I have ended up with alternative realizations at the level of dark nucleons and sequences of 3 dark nucleons [**?**].

3.1 Summary of Combinatorial Hierarchy

I summarize first the basics of CH.

1. One considers the space algebra of Boolean statements of n bits which can be also extended to complex linear space -quantum Boolean algebra. One can give it linear structure as Z_2 algebra for binary coefficients with Z_2 sum having set theoretic interpretation. This linear space has some basis. That the coefficient field for linear structure is Z_2 does not seem to be absolutely essential. In TGD framework one considers the linear space defined by quantum Boolean algebra with qubit interpretation generated by fermionic oscillator operators: one operator for every bit.

2. One assigns to the linear n-D space the n^2-D space of linear maps of it to itself. One can also consider the space of maps of quantum Boolean algebra to itself and also require that this defines a Boolean homomorphism. Dimensions would be the same: only coefficient field would be different.

3. To CH level, which corresponds to Mersenne prime $M(n) = M_{M(n-1)}$ ($n = 2, 3, 7, 127, 2^{127} - 1, ...$) one assigns vector space with dimension

$$D(n-1) = [(M(n-1)+1)]^2 \ ,$$

and requires that the space formed by

$$D_1(n) = \frac{(M(n)+1))}{2}$$

bit sequences, which represent a subset of mutually consistent Boolean statements as subset of $M(n)+1$ bit sequences are representable as a subset of bit sequences with $D(n-1)$ bits. This demands

$$D_1(n) \leq D(n-1)$$

giving

$$\frac{M(n)+1)}{2} \leq [M(n-1)+1]^2 \ .$$

4. This criterion is satisfied for the primes of CH up to M_7 but not for M_{127}: $2^{127} - 1 > 128^2$ so that M_{127} should not included if I have understood the criterion correctly.

   For $M_7 = 2^7 - 1 = 127$ one obtains the condition $2^6 = 64 \leq 8 \times 8 = 64$ so that condition is saturated. Remarkably, 64 is the number of DNA codons!

5. The numbers of CH are also known as Catalan Mersenne numbers. Catalan Mersenne primes are special case of double Mersenne primes M_{M_n} (see `http://tinyurl.com/j4tqwch`). Catalan conjecture that Catalan Mersennes are primes up to some limit. After the first non-prime the remaining Catalan Mersenne numbers are necessarily composite. The known double Mersennes are given by M_{M_p}: $p = 2, 3, 5, 7$. No other cases are known. These primes are good candidates for labelling scaled up variants of say hadron physics. To my opinion Catalan criterion is more plausible.

6. Classical number fields are in key role in TGD [K24, K15, K20] and have dimensions $D = 1, 2, 4, 8$. Also CH involves these dimensions. $D(n-1) = M(n-1)+1$ giving dimensions $2, 4, 8$ for M_2, M_3, M_7. For M_{127} one would obtain $D = 128$, which does not correspond to any division algebra. This might relate to the above observation.

3.2 CH as a prediction of quantum TGD

In the following the interpretation of Boolean map in ZEO is proposed. Also it is shown that M_7 level allows a natural realization in terms of spin-isospin states of fermions and that M_{127} level is obtained in second quantization meaning going from the level of imbedding space to the level of WCW.

3.2.1 Interpretation of the lower level Boolean map in terms of ZEO

One can ask, why one should have this kind of map? One interpretation is that the space of Boolean statements at given level is imbeddable to the space of quantum Boolean maps at previous level. Quantum Boolean maps would represent Boolean rules A→ B, "theorems" or "laws of physics".

1. In TGD framework the interpretation of CH would be as a hierarchy of statements about statements about... The number of statements about N statements is indeed 2^N. One statement corresponding to all bits equal to 0 (in set theoretic realization empty set) is thrown away so that one has $2^N - 1$ statements instead of 2^N.

2. ZEO means that physical states are pairs of states with opposite conserved quantum numbers: they correspond to physical events, which replace states as fundamental entities in ZEO. The fermionic parts of positive and negative energy parts of states would be pairs of many-fermion states allowing interpretation as elements of quantum Boolean algebra. Zero energy states themselves would correspond to pairs of these fermionic states and thus to "theorems" A→ B or maps from Boolean algebra to itself. The allowed statement pairs would satisfy fermion number conservation and conservation of various quantum numbers and would indeed represent laws of physics.

3. A possible interpretation of the map would be that the statements at given level $M(n+1)$ must be representable as theorems at previous level $M(n)$. For $M(n) > M_7 = 127$ this would not hold true anymore. Could this have some deep mathematical meaning as the wild association with Goedel's theorem suggests?

   In the model of genetic code and its generalizations [K5] I have proposed that each level of CH defines a maximal number of set of statements forming a set theoretic inclusion hierarchy and logical implication hierarchy: the number is 2^{n-1} for 2^n n-bit statements. For $M_7 = 127$ the number is 64, the number of DNA codons, which would thus have interpretation as axioms or "fundamental truths". In this case the representability would still hold and map would be bijection. At the next level one would have "memetic code" with 2^{126} codons representable as sequences of 21 DNA codons with stop codon included ($126 = 21 \times 6$). By the proposed criterion, at memetic level only vanishingly small subset of truths would be representable as theorems at genetic level.

3.2.2 Representation of M_7 level in TGD framework

Could the saturation for M_7 have some physical meaning? The maps would be from 8-D space to itself.

1. Bits can be represented in terms of spin and electroweak spin giving $2 \times 2 = 4$ states and imbedding space-spinors ($H = M^4 \times CP_2$) of given H-chirality (quark or lepton like), given fermion number (fermion or antifermion) and physical helicity. If also unphysical helicities with fixed fermion number are allowed one would have $4 + 4 = 8$ states. The condition that helicity is physical would reduce the number of states by one half. This applies to both quarks and leptons since color is not spin like quantum number in TGD (colored states correspond to partial waves in CP_2).

2. What could be the interpretation for $2^7 - 1 = 127$ states containing as subset $n = 2^6$ states. Could $n = 2^6$ correspond to the number of states in the tensor product formed by pairs of 8 leptons and 8 antileptons allowed to have also unphysical polarizations? Same would apply to quarks. Allowing both quark-antiquark and lepton-antilepton type states one would have 128 states. The physicality condition for boson polarizations could drop the number of states

to 64. What the dropping of one state would correspond: to the dropping of $\nu_R - \overline{\nu}_R$ pair having no electroweak and color couplings perhaps?

One can imagine two alternative identifications for the two tensor factors.

(a) In TGD framework fundamental bosons correspond to fermion antifermion pairs with members at opposite throats of wormhole contact connecting two space-time sheets. Could the genetic code correspond to 64 elementary bosons with physical polarizations and the maps to those assigning to 8 fermions 8 antifermions?

(b) An alternative identification is suggested by ZEO. The tensor product of fermionic Boolean algebras at opposite boundaries of causal diamond (CD) would replace that at opposite wormhole throats. This would in accordance with the interpretation of zero energy states as statements $A \to B$ represented as Boolean maps.

3.2.3 Representation of M_{127} level in TGD framework

What about the physical interpretation of M_{127} level in TGD framework?

1. The first thing to observe is that physically $p = M_{127}$ corresponds in TGD to the p-adic prime p characterizing electron in p-adic mass calculations: Compton length is proportional to the p-adic length scale and thus proportional to $\sqrt{p}$. The remaining Mersenne primes correspond to completely super-astrophysical Compton lengths. Hence M_{127} has a very special role. The Mersenne primes $3, 7, 31, 127$ giving rise to double Mersenne primes correspond to extremely short p-adic length scales.

   Recall that the ratio of m_{CP_2}/m_e is approximately $m_{CP_2}/m_e = 2^{127/2}/\sqrt{5+x}$, where $x \in [0,1]$ characterizes the second order contribution to electron mass from p-adic mass calculations [K9]. The ratio of Planck mass to proton mass equals to $m_{Pl}/m_p = 1.307 \times 10^{19}$. For $x = 0$ this gives $m_{Pl}/m_{CP_2} = (m_p/m_e) \times 3.96 = 7.271 \times 10^3$, which is not far from $2^{13} \simeq 8.912 \times 10^3$. The value of 2^{13} is very attractive number theoretically and would be obtained for $x = .5$, again power of 2.

2. The states at this level should correspond to statements about statements at the lower level represented in terms of quark lepton state space as many-fermion states assignable to wormhole throat or several wormhole throats (elementary corresponds to two wormhole contants and 4 wormhole throats). The construction of infinite primes can be interpreted as a process of forming repeatedly statements about statements and the physical analog is repeated second quantization [K23].

   In the recent situation second quantization would correspond to the formation of many-fermion states at partonic 2-surfaces defined by the throats of wormhole contacts. This would automatically give rise to M_{127} states if one has 127 single fermion states to begin with.

   Physically this step would correspond to a step from the spinor modes of imbedding space to the spinor modes of WCW identifiable as fermionic Fock states assignable to partonic 2-surfaces so that indeed a huge abstraction is in question. I have proposed that anyonic states could be this kind of states for large value of $h_{eff} = n \times h$ implying that the size of wormhole throat becomes nano-scopic [K22].

3. One has 127 boson states but how to obtain 127 (or $128 = 2^7$) single fermion states? Counting only spin and weak isospin gives $n = 8 + 8 = 2^4$ ($n = 4 + 4 = 2^3$) single fermion states if one allows (does not allow) also unphysical polarizations. The simplest option is that each single fermion state has 2^3 (2^4) additional states. The location of fermion at one of the 4 wormhole throat could give 4 additional degrees of freedom. This would leave 2 (4) additional states per fermion state still missing.

4. A good guess is that quark color realized as color partial waves comes in rescue and gives the needed states. Light quarks must move in color triplet states and leptons in singlet states. Thefore quarks have $3 \times 8 = 24$ modes and leptons 8 modes giving altogether 32 modes altogether. There are 4 wormhole throats so that $4 \times 32 = 128$ modes are obtained

and if right-handed neutrino is thrown out one has 127 states as required if no constraints on polarizations are posed. It therefore seems that TGD physics codes CH naturally at elementary particle level!

There is indeed a rich set of "vibrational" degrees of freedom giving also rise to color degrees of freedom. The symplectic group of $\Delta M^4_{\pm}$ assignable to either boundary of causal diamond (CD) defined as the intersection of future and past directed light-cones of M^4 with points replaced with CP_2 gives rise to products of S^2 and CP_2 partial waves. Besides this there is a conformal weight labelling the states correlating with $S^2 \times CP_2$ partial wave Light quarks massless before massivation by p-adic thermodynamics move in color partial waves and color triplets are obtained as the color excitations for them corresponding to higher conformal weights and having CP_2 mass as mass scale.

I have already earlier ended up with the proposal that genetic code is realized at the level of dark nuclear physics. Either the states of dark proton or sequence of 3 protons could be organized naturally states corresponding to 64 DNAs, 64 RNAs, 20 aminoacids, and 40 tRNAs and vertebrate genetic code follows from very simple assumption that opposite spins are paired [K21, K6] [**?**] (see `http://tinyurl.com/jgfjlbe`). These findings suggest that genetic code and memetic code are also realized at the elementary particle level.

Acknowledgements: I am gratetul for James Bowery for raising the question about the possible relevance of CH for TGD.

4 Number theoretical models for genetic codes

The naïve thinking would suggest that the DNA-amino-acid correspondence is unique and same in the alien biology as in our biology. This is not the case. The notion N-particle leads to a model how N-hydrogen atoms define names for molecules and how molecules with conjugate names form especially stable bound states and how the same mechanism explains lock and key mechanism of bio-catalysis. The lock and key mechanism depends only weakly on chemistry and it is quite possible that several genetic codes are realized.

Hence the tRNA molecules mediating DNA-amino-acid correspondence could be different for various life-forms. The stability of various possible tRNA type molecules determining the code would be determined by the electromagnetic environment. Therefore one must take genetic code as a result of selection. The findings about the alien codes, if taken seriously, suggest also guesses about the origin of the genetic code.

The basic new result inspired by the attempt to identify the alien genetic code is the finding that both our and alien genetic codes factorize in a good approximation to a product codes associated with DNA doublets and singlets. This raises the question whether the factorization occurs also at the level of amino-acids. Could DNAs triplets have resulted as a symbiosis of singlets and doublets whereas amino-acids might have been developed via a symbiosis of 2 (3) molecules coded by 4 DNA singlets and 10 (7) molecules coded by 16 DNA doublets?

4.1 Three kinds of number theoretical models for the genetic code

TGD has led to three different number theoretic approaches concerning the understanding of the genetic code.

1. In [K5] the model of the genetic code based on the notion of Combinatorial Hierarchy is discussed. This approaches predicts at least one additional code that I have christened memetic code.

2. In [K17] a universal number theoretical code giving genetic code as a special case and based on the maximization of a number theoretic information measure was developed.

3. The model based on the assumption that genetic code has evolved from a product code is the one to be discussed in this chapter (see also the discussion in [K19]).

4.1.1 Genetic codes as deformations of product codes

In this section number theoretical models based on the approximate factorization of the genetic code into product code formed by doublet and singlet codes are 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 labelled by the first two bases of DNA triplet. One can invent an information maximization principle providing a quantitative formulation for this tendency.

4.1.2 Genetic codes based on the maximization of number theoretic information measure

In the chapter [K17] an alternative number theoretic model for the ordinary genetic code and its variants is discussed. This model is based on very general number theoretic notions, in particular, number theoretical generalization of Shannon entropy, and must be regarded as the most convincing one of the three number theoretic models constructed hitherto. This model allows to identify ordinary genetic code and its variants as codes maximizing a unique number theoretic information measure. The model is also consistent with the idea that genetic code has evolved from a product of singlet and doublet codes.

The model predicts the number for "amino-acids" once the number n of "DNAs" is known as $N(n) + 2$, where $N(n)$ is the number of primes not larger than n. For 80 DNA triplets the prediction would be $24 = 3 \times 8$ rather than 23 amino-acids. Hence the two models for the genetic code would not be consistent.

Before making any hasty conclusions one should recall that the interpretation of the Crabwood circle as ASCII text involves considerable uncertainties. A modification of single special symbol or small letter to a symbol not appearing in the proposed interpretation of the Crabwood message would give 24 "amino-acids". For instance, the ASCII symbols for dot *resp.* comma are 00110100 *resp.* 01110100 and differ only by a single bit so that misinterpretation cannot be excluded.

This model of genetic code emerged much later than the model for alien genetic codes and is not discussed in this chapter.

4.2 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.

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. Indeed, the identification of pre-amino-acids as DNA singlets or doublets dictated by RNA-DNA translation mechanism is strongly favored by the physical model for the evolution of the genetic code. With this identification triplet pre-amino-acids (DNA triplets) are simply composites of doublet and singlet pre-amino-acids (DNA doublets and singlets).

Despite this interpretation, the study of the amino-acid geometric structure is in order. It 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.

1. 2×10 *product structure at amino-acid level*

2×10 decomposition for real amino-acids might approximately correspond to hydrophobic-hydrophilic dichotomy which plays a key role in the amino-acid chemistry. This correspondence cannot be very precise since the number of the hydrophobic (-philic) amino-acids is 8 (12) rather than 10 (10). Of course, this is what one expects since the product symmetry is broken.

2. 3×7 *product structure at amino-acid level*

Aminocids can be classified into three groups. The first class contains 8 hydrophobic non-polar amino-acids: ala, val, leu, ile, pro, met, phe, trp, Second class consists of 7 hydrophilic polar amino-acids gly, ser, thr, cys, asp, glu, tyr. The third class consists of polar hydrophilic acidic amino-acids asp, glu and hydrophilic basic amino-acids lys, arg, his: 5 altogether.

Could these three classes correspond to the 3×7-ness?

1. First of all, the non-varying group contains almost(!) as a rule both the acidic carboxy group $COOH$ which tends to ionize to COO^- and basic aminegroup NH_3 which tends to ionize to NH_3^+. When carboxy or amine group is associated with the side group, the 2+3=5 acidic or basic polar amino-acids result. Thus the three-ness in standard sense corresponds to the difference for the total numbers of acidic and basic groups of the side chains: amino-acid side chain is either neutral and non-polar, neutral and polar, or charged. This leads to 8+7+5 decomposition and a slight breaking of three-ness.

2. One could however consider a modified definition in which one counts the numbers N_+ of basic and N_- of acidic groups of the *entire* amino-acid and uses the difference $N_+ - N_-$ to tell the net charge of the amino-acid. If this criterion is used, the first group contains one alien, proline. Proline differs from all other amino-acids in that the neutral group $H_3N^+ - COO^- - C - H$ group is replaced by a charged $HN - COO^- - C - H$ group. But this means nothing but replacing the basic group NH_3^+ with a non-basic NH. This implies also a net charge for proline. If net charge is taken as the characterizing property of the third group of amino-acids, proline belongs to it. Therefore first and second would group contain 7 amino-acids and the third group would contain 3 positively charged and 3 negatively charged amino-acids.

3. If one thinks that stopping sign formally corresponds to one additional amino-acid in the third group, one indeed has 7+7+7 decomposition. For some rare life-forms to be discussed later stopping sign codon ATC can code for both stopping sign and non-standard amino-acid pyrrolysine depending on context [I7]. Pyrrolysine, being a derivative of lysine, is basic so that in this case one would have 7+7+7 decomposition even without counting stopping sign formally as an amino-acid.

The 7-ness index labelling the amino-acids with the three groups should be some abstract property and it is impossible to make any conclusions on basis of the chemical formulae alone.

3. Is the product structure at the level of amino-acids really needed?

It has become clear that the product structure for amino-acids is not necessary.

1. The number theoretic model of the genetic code discussed in [K17] neither predicts nor requires the product structure for amino-acids but is consistent with the approximate product structure for codons.

2. In [K19] a model for the evolution of the genetic code from a product code mapping RNAs to a subset of RNAs is studied. In this model the product structure at the level of coded RNAs is natural but there is no reason for it at the level of amino-acids which, according to the model, originally only catalyzed RNA→ RNA mapping but later replaced the coded RNAs in a kind of palace revolution.

4.3 Number theoretical model for the terrestrial genetic code

The study of the terrestrial genetic code allows to deduce the process leading to the breaking of the product symmetry and T-C symmetry. This process turns out to work as such also in case of alien codes.

4.3.1 Approximate reduction to a product code

The dependence of the amino-acid coded by DNA on the third codon of DNA triplet is weak and Crabwood message suggests that both doublet and triplet codes are realized. 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. There are two options.

1. 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.

2. The decomposition $21 = 3 \times 7$ for amino-acids plus stopping sign suggests that singlet code maps four bases to 3 "pre-amino-acids" and 16 doublets to 7 "pre-amino-acids". In the first approximation the triplet code would decompose to a product of doublet code and singlet code in the sense that 4 singlets are mapped to Z_3 and 16 doublets are mapped to Z_7 so that 21 different product states result. The decomposition of the statements consistent with some atomic statements suggests itself strongly. In the first approximation the triplet code would decompose to a product of doublet code and singlet code in the sense that 4 singlets are mapped to Z_3 and 16 doublets are mapped to Z_7 so that 21 different product states result. The problem of this option is that it predicts complete breaking of T-C symmetry and the breaking of the product symmetry should produce T-C symmetry. This looks two complicated.

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.

In 3×7 case singlet code would have following interpretation. Z_3 is identified as negations of 4 selected statements with 00 excluded. Statement and its negation are projected to this Z_3 representing negations with 00 excluded so that 11 must be projected to some other statement. The degeneracies of the code are unique: 2, 1, 1 since any change of the code changing this degeneracy spectrum implies that one degeneracy vanishes.

Same applies to Z_7 and 16 DNA doublets. Now 1111 is mapped to some statement in the set of negations. In this case the simplest coding is obtained by mapping 7 statements to their conjugates and the two remaining statements to different conjugate statements in the set of 7 statements. The resulting degeneracy structure is 2222233 and entropy is maximal for this code.

4.3.2 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.

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

Table 6: 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.

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.

4.3.3 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 labeled by first and second bases of DNA triplet should have similar amino-acid structure. For 3×7 code the 4-column should have $AABC$ structure. This is not case. Almost all 4-columns have $AABB$ structure and there are also many $AAAA$ type 4-columns. 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 3×7 code a given amino-acid should be coded by DNA pairs of form (XYA, XYG), or DNA of form XYC or XYT. 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. For 3×7 each 2n-plet formed by degenerate (XYA, XYG)-pairs is accompanied by n-plets of type XYT and XYC. In 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 codes 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.

The **Table 9** 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 7** 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.

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.

	A	G	T	C	
A	phe	ser	tyr	cys	A
	phe	ser	tyr	cys	G
	leu	thr	stop	thr	T
	leu	thr	stop	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	asn	ser	A
	ile	ser	asn	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 7: 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 8: 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 9: The code table after the T-C symmetry breaking

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) sidegroups. 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. As will be found trp is also otherwise singular case.

A working hypothesis worth of studying is that the symmetry breaking mechanism is universal and applies also to the capital letter code and even to the small letter + special symbol code in an appropriately generalized form. This hypothesis is highly predictive, and the fact that one can produce these codes using the product ansatz, the same "volume preserving flow", and T-C symmetry breaking, encourages to think that the picture has some truth in it.

4.3.4 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.

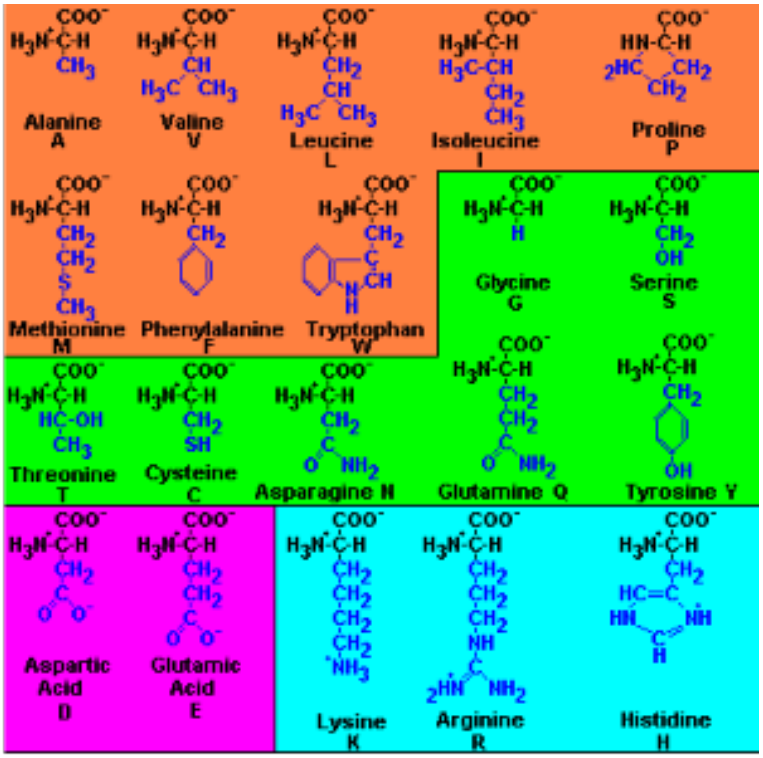


Figure 3: The chemical structure of amino-acids. The first group (ala, ...) corresponds to non-polar amino-acid side groups, the remaining amino-acids to polar side groups. The two lowest groups correspond to acidic (asp, glu) and basic side groups.

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.

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-acidss 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

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

and to the entire table the probability

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

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

$$\sum_k p_k = 1 \quad .$$

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

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

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

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

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

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

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.

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

The deviations of the terrestrial genetic code from the standard code [I7] 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. The construction of the small letter code indeed forces to assume this kind of breaking of universality. 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 [I7] 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. This suggests that also in the case of the capital letter code the amino-acid coded 8 times is thr. In case of the small letter + special sign code the 13-fold degerate amino-acid could be thr.

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 imbedding 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.

   The small variants of the letters K and V are lacking from small letter+special sign code. This might signal that the corresponding amino-acids are replaced by selenocystein and pyrrolysine represented by h and \ in the small letter code.

4.4 Capital letter code as a product code with broken T-C symmetry

What about capital letter code: does it also have approximate product structure? Product structure predicts that many degeneracies, in particular the largest degeneracies should be divisible by two. In case of 2×10 code all degeneracies are predicted to be divisible by two. This is not the case now as **Table 9** shows. One can however try to find a product code which is as near as possible to the real one.

The degeneracies 1111111234 for the doublet 2×10 representation differs from our genetic code in that 1111112233 is modified to 1111111234. These degeneracies would be the degeneracies most naturally associated with the 16 DNA doublet code with 10 "pre-amino-acids" possibly associated with plasmoid like life forms serving as messengers of the aliens.

The simplest option would be that this correspond to taking one doublet from second 2 and adding it to second 3 so that one additional singlet results. Unfortunately, the fact that stopping sign has degeneracy 7(8) excludes this option.

The 1111111234 decomposition predicts the following numbers for DNAs with various degeneracies. Also the corresponding numbers for capital letter code are included.

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

Table 10: 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.

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

Table 11: Capital letter code table before the flow

The following process gives the degeneracies of the OPpose code.

1. Take one DNA from second 8-plet and add it to 6-plet to get two 7-plets so that one has $N(7) = 2$ and $N(6) = 1 > 0$.

2. Change one DNA in 6-plet to the DNA which does not exist in the table to get $N(6) = 0$, $N(5) = 1$, $N(1) = 1$. The non-existing DNA is generated in essentially the same manner also in case of our code.

3. One can transform 7 2-plets into 2 3-plets, 4-plet and 4 singlets as follows. Take from two doublets one DNA and move them to third doublet to get $N(1) = 3 < 5$, $N(2) = 11 > 7$, and $N(4) = 3$. There are four superfluous doublets remaining and forming pairs. For each pair take DNA from one doublet and move it to second one to get $N(1) = 5$, $N(2) = 7$ and $N(3) = 2$.

Assuming that the decomposition of DNA doublets is obtained from that for our code in the proposed manner and that the same flow induces T-C symmetric part of the breaking of the product symmetry, one can fix the DNA-amino-acid correspondence highly uniquely for the capital latter code. The unbroken code contains two octets. Since for yeast mitochondria both GA and TA columns code for thr, the guess is that the second octet corresponds to thr. The second octet must be ser from the product symmetry. The requirement that the code table resembles as much as possible the code table of our genetic code leads to the following working hypothesis for the code table before symmetry breaking.

T-C symmetry breaking can be understood as follows.

1. Take one DNA from second 8-plet (ser or thr and add it to 6-plet representing stopping sign

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

Table 12: Capital letter code table after the flow

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

Table 13: Capital letter code table after the T-C symmetry breaking

to get two 7-plets so that one has $N(7) = 2$ and $N(6) = 1 > 0$. Thr is chosen in the sequel for definiteness and corresponds to TGC.

2. Change one DNA in thr 6-plet to the DNA which does not exist in the table to get $N(6) = 0$, $N(5) = 1$, $N(1) = 1$. The non-existing DNA is generated in essentially the same manner also in case of our code. stop at ACT is transformed to trp as so that trp is in the same position as in our genetic code.

3. What one must do is to transform 7 2-plets into 2 3-plets, 4-plet and 4 singlets. This is achieved in the following manner.

   (a) Take from two T-C doublets one DNA and move them to a third doublet to get $N(1) = 3 < 5$, $N(2) = 11 > 7$, and $N(4) = 3$. For instance, this is achieved by transforming glu and ala to asp. The value of information measure decreases by $log(64/27)$ in this process. There are also many other manners to do this.

   (b) There are four superfluous doublets remaining and forming pairs. For each pair take DNA from one doublet and move it to second one to get $N(1) = 5$, $N(2) = 7$ and $N(3) = 2$. More concretely $(AA)_2$ leu doublet is transformed to phe-leu, and $(TA)_2$ met-doublet is transformed to ile-met so that correct degeneracies result and the information measure increases in these processes by $2 \times log(27/16)$ which is larger than $log(64/27)$ so that the net increase of the information measure is positive in the entire process.

The process is not obviously completely unique but the proposed choice is favored because the small latter+special sign code can be obtained as a small deformation of this code.

4.5 T-C symmetric models for small letter plus special symbol code

One can apply T-C symmetric product model with symmetry breaking also to the code candidates involving small letters. There are three candidates for these codes.

1. The 4×17 code with 18 amino-acids involving only small letters with h interpreted as stoppping sign: this code makes sense for Oppose option only and since the expressive power is not maximal, it will not be discussed in the sequel.

2. $4 \times (16 + 4)$ code with 23 generalized amino-acids (\, h, and special symbols $!, \&, .$ are interpreted as belonging to the extended family of amino-acids).

3. The $4 \times (16 + 4)$ code with 20 amino-acids (\ and h are interpreted now as amino-acids). This code results from the code with 23 generalized amino-acids by assuming that the DNAs coding for $!, \&$ and period code for the stopping sign.

The candidates 2) and 3) appear as Oppose and OPpose options.

4.5.1 The nature of silicon modification

The product model for the genetic codes suggests an interpretation of the small letter codes. The Chilbolton message tells that also silicon is fundamental for the alien life at DNA level so that one can consider the possibility that one of the DNA and RNA doublets is modified by an addition of something containing silicon to give an additional doublet.

For $(4 + 16) \times 4$ code four additional doublets must be present. If some base of DNA suffers a modification, it suffers the modification also if it appears in RNA triplet at the same position, and this in turn implies that also the conjugate of the DNA base suffers modification so that 32 additional triplets are generated. Thus the modified base of DNA cannot appear in RNA and vice versa. DNA bases (A, G, T, C) correspond to RNA bases (U, C, A, G). Since the T of DNA corresponds to the U of RNA, there is only one possibility. The modified base is T for DNA and U for RNA, and the T_S of DNA must correspond to U_S of RNA rather than A_S. The simplest possibility is that the doublets of form XT have doubled by the silicon modification of the second T to XT_S. Also T_SX type modification is in principle possible but the construction of the code favors the XT_S option (in this case code the table gets a fifth column whereas for T_SX gives rise to a fifth row).

n	1	2	3	4	5	6	8	9	10	12	13
N	0	16	0	4	0	2	0	0	2	0	0
N	10(9)	4(5)	0	3	2(3)	3(2)	0	1	0	0	1

Table 14: The numbers $N(n)$ of amino-acids coded by n DNAs for code containing small letters and special symbols for 2×12 option. Both OPpose and Oppose options are included.

4.5.2 2×12 product model for the small letter plus special symbol code with 80 generalized DNAs and 23 amino-acids

The optimal candidate for the code involving 64+16 generalized DNAs involves 20+3 generalized amino-acids. There are two options corresponding to the decompositions $24 = 3 \times 8$ and $24 = 2 \times 12$. The assumption that small letter plus special sign code follows from the capital letter code as extension favors 2×12 option. 2×12 option for the small letter + special sign code allows highly unique model since one can assume that the code results as a simple extension of the capital letter code and is obtained by the same symmetry breaking procedure as the capital letter code and terrestrial genetic codes. The discussion below is restricted to OPpose option.

The first step is to deduce the composition in the set of $4 + 16$ DNA doublets defining the product code. The only working option has the decomposition 111111112235, which corresponds to the decomposition

$$20 \times (2 \oplus 2) = (5 \oplus 3 \oplus 2 \oplus 2 \oplus 8 \times 1) \times (2 \oplus 2) \ .$$

This gives **Table 14** for the degeneracies.

The breaking of the product symmetry looks large but it turns out that the code can be obtained as a relatively small deformation and extension of the capital letter code.

The first things to observe about the code are following.

1. Comparing the decomposition 111111112235 with the corresponding decomposition 1111111234 for the capital letter code, one can guess that the small letter code is obtained from the capital letter code by the following process in the set of 4 exotic RNA-doublets. Decompose the four exotic RNAs to $(2 \oplus 1 \oplus 1) \times (2 \oplus 2)$ such that $2 \times (2 \oplus 2)$ codes for exotic and ordinary amino-acid quartet. Since trp is lacking from capital letter code before symmetry breaking, one can assume that trp is the ordinary amino-acid. Since the exotic amino-acid "period" appears five times, the second 4-plet must code for "period". The two doublets must code for exotic doublets & and ! which reduce to singlets after symmetry breaking. Two exotic doublets fuse with the two octets of the capital letter code to code for two decouplets and must therefore code for the ordinary amino-acids ser and thr. Thus the code table without symmetry breaking looks very much like capital letter code table.

2. The modification $XT \to XT_S$ implies that code table gets fifth column. Only this option allows to generalize in non-trivial manner the flow and allows to see trp 4-plets as being consistent with product code.

3. Terrestrial codes contain two exotic amino-acids scys and plys. The fact that the small letter + special sign code contains the symbols h and \ with ASCII number larger than 64 not appearing in the capital letter code is taken as a suggestion that the corresponding amino-acids are exotic. A natural working hypothesis is cys is replaced with scys and lys with plys. Needless to add, this hypothesis must be taken with a grain of salt.

1. Product code before flow

The code table before the action of the flow and T-C symmetry breaking looks like follows. The code table obviously resembles capital letter code table to a very high degree and satisfies all the constraints resulting from the A-G and T-C symmetries and product structure of the code.

2. The action of the flow

	A	G	T	C	T_S	
A	phe	ser	tyr	scys	.	A
	phe	ser	tyr	scys	.	G
	leu	thr	stop	stop	thr	T
	leu	thr	stop	stop	thr	C
G	val	ala	glu	gly	!	T
	val	ala	glu	gly	!	C
	thr	stop	gln	arg	trp	T
	thr	stop	gln	arg	trp	C
T	ile	ser	asn	ser	.	A
	ile	ser	asn	ser	.	G
	met	thr	plys	arg	&	T
	met	thr	plys	arg	&	C
C	val	ser	asp	scys	ser	A
	val	ser	asp	scys	ser	G
	thr	pro	his	arg	trp	A
	thr	pro	his	arg	trp	G

Table 15: Small letter special sign product code before flow and T-C symmetry breaking.

	A	G	T	C	T_S	
A	phe	ser	tyr	scys	.	A
	phe	ser	tyr	scys	.	G
	leu	ser	stop	stop	thr	T
	leu	ser	stop	stop	thr	C
G	thr	pro	his	arg	trp	A
	thr	pro	his	arg	trp	G
	thr	stop	gln	arg	trp	T
	thr	stop	gln	arg	trp	C
T	ile	thr	asn	ser	.	A
	ile	thr	asn	ser	.	G
	met	thr	plys	arg	&	T
	met	thr	plys	arg	&	C
C	val	ser	asp	scys	ser	A
	val	ser	asp	scys	ser	G
	val	ala	glu	gly	!	T
	val	ala	glu	gly	!	C

Table 16: Small-letter special sign genetic code after the flow and before T-C symmetry breaking.

	A	G	T	C	T_S	
A	phe	ser	tyr	cys→ scys	.	A
	phe	ser	tyr	cys→scys	.	G
	phe	ser	stop	stop→trp	thr	T
	leu	ser	stop	trp	thr	C
G	thr	pro	his	arg	trp	A
	thr	pro	his	arg	trp	G
	thr	stop	gln→phe	arg	trp	T
	thr	stop	gln	arg	trp	C
T	ile	thr	asn	ser	.	A
	ile	thr	asn	ser	.	G
	ile	thr	lys→ ile	arg	.	T
	met	stop	lys→ plys	arg	&	C
C	val→ ser	ser	asp	gly→scys	ser	A
	val→ser	ser	asp	gly→scys	ser	G
	val→ ser	asp	asp	gly→asp	asp	T
	val	ala	glu	gly	!	C

Table 17: Small letter special sign genetic code resulting from T-C symmetry breaking. The replacements $X \to Y$ tell how the code in the sector of ordinary DNAs is obtained from the capital letter code.

3. T-C symmetry breaking

The basic assumptions are that the G-column of the code is universal for the alien code just as it is universal for the terrestrial codes, and that the code table resembles maximally to our code table and capital letter code table.

1. One must transform the two 10s (thr and ser) to 13 and 9. The clue to the symmetry breaking mechanism comes from the finding that one must be able to generate as many as 10 singlets. Hard trial and error work teaches that one cannot get these singlets unless one allows $10 + 4 \to 13 + 1$ mechanism for producing one of the singlets. The transformation of val-val-val-val to ser-ser-ser-val is the only candidate for this transformation and gives $N(4) = 3$ (scys, period, trp) and $N(1) = 1$.

   The thr is the second 10-plet and the transformation of TTC-thr to stop is the only possibility if the universality of the G column in alien sector is assumed. The transformation of $(AC)_2$ stop-stop column to trp-trp implies maximal resemblance with our genetic code, and one obtains $N(13) = N(9) = 1$ (thr, ser), $N(6) = 2$ (arg, trp), $N(5) = 1 < 5$ (stop) and $N(4) = 2 < 3$ (scys, period).

2. The remaining transformations must produce $N(1) = 10 > 1$, $N(2) = 4$, $N(4) = 3 > 2$, $N(5) = 2 > 1$, $N(6) = 3 > 2$ by acting on the T-C type doublets only and thus generating a breaking of T-C symmetry. The first step is to replace & in the $(TT_S)_2$ by "period" to get $N(5) = 2$, $N(4) = 1$, $N(1) = 2$. What one must create by the splitting all the remaining T-C doublets so that 2 4-plets and 1 6-plet as extension of A-G type doublets results. The choice of the A-G type doublets is not unique but the requirement that the code table resembles maximally the code table of the capital letter code fixes the choice of A-G type doublets extended to 4-plets to be AA_1 (phe), $(TT)_1$ (ile) and the A-G type doublet extend to 6-plet to be CT_1 (asp). **Table 17** summarizes one possible code table satisfying these constraints. For comparison also the table for capital letter code is given.

4.5.3 Product model for the small letter code with 20 amino-acids and 80 generalized DNAs

The number theoretical model generalizes for the codes defined by 64 ordinary DNAs + 16 DNAs of form XT_SY and assuming that besides 20 amino-acids there are 3 additional modified amino-acids. A small letter-special symbol code with 80 DNAs and 20 amino-acids is obtained from 23-amino-acid code by assuming that the exotic DNAs coding for special signs $!, \&$ and period code for stopping sign and the previous construction for 2×12 code works as such. Oppose option with 64 DNAs (special signs being not interpreted as belonging to the code) and 18 amino-acids is in conflict with the requirement of a maximal expressive power. My personal conviction is that this option can be safely forgotten.

4.5.4 Why the numbers 64 and 80?

The dark matter hierarchy based on the hierarchy of increasing values of Planck constant predicts that the entire universe is a macroscopic quantum system and elementary particles have a hierarchy of zoomed up variants with arbitrarily large Compton length (proportional to $\hbar$) [K18]. Dark matter should be especially important for living matter and life should therefore involve fundamental physics in an essential manner rather than emerge at some very high level of complexity. Hence one can ask whether the numbers 64 and 80 for the codons of the two codes could reflect basic facts about fundamental physics in TGD Universe. The following numerological argument based on detailed counting of particle states encourages to take this idea half-seriously at least.

1. Gravitons and more general stringy states are not counted since they correspond to bound states of fermions and bosons connected by flux tubes. Color is counted neither since it corresponds to CP_2 partial wave and is not spin like degree of freedom in TGD framework. Family replication phenomenon has a topological explanation and is counted neither. This leaves only spinorial degrees of freedom which according to TGD inspired theory of consciousness are responsible for Boolean representations using fermionic Fock states. The natural guess is that these fermionic degrees of freedom might relate to the genetic code or genetic code might represent them.

2. TGD predicts in purely spinorial degrees of freedom 8 lepton states (lepton and anti-lepton both having 4 states due to spin and electro-weak isospin). Also phase conjugates of these states are predicted so that 8+8=16 states are obtained. The number of spinor states is same in the quark sector. This gives 16+16=32 states altogether.

3. Bosons are identifiable as tiny wormhole contacts carrying fermion and anti-fermion numbers at the light-like wormhole throats. Essentially lepto-antilepton and quark-antiquark pairs or their superpositions are in question. $(2 + 1) \times (3 + 1) = 12$ leptonic and 12 quark like bosons with spin and electro-weak isospin equal to 1 or 0 (only two massless spin states are possible). Together with phase conjugates this makes 24+24= 48 states. 24 of them correspond to ordinary electro-weak gauge bosons and Higgs and the remaining 24 are exotic bosons with charge matrices orthogonal to the charge matrices of electro-weak gauge bosons. For exotic counterparts of W bosons and Higgs the sign of the coupling to quarks is opposite. For photon and Z^0 also the relative magnitudes of the couplings to quarks much change. The total number of bosonic states is 48 and the number of all particle states in this sense is 48+16+16=80. If quarks are dropped from consideration the number is 64.

4. The numerological question is whether the 64 ordinary genetic codons are in some deeper sense in one-one correspondence with 48 color singlet gauge bosons and 16 lepton states and the 80 codons of the extended code in one-one correspondence with all states constructed in this manner.

4.6 Imbedding of the amino-acid space into DNA space and the universal part of the genetic code

The concrete geometric formulation for the symmetries is based on the imbedding of 20+1 generalized amino-acids to the space of 64 DNAs. Obviously, the amino-acids are coded by the DNAs to

which they are mapped by this imbedding. There is indeed an imbedding of 20 amino-acids plus stopping sign with 2×10 structure to the set of 64 DNA triplets which have 4×16 structure. 2 is imbedded into 4 which corresponds to the 4 last bases of DNA and 10 into 16 which corresponds to 16 pairs of first two bases of DNA. The lacking amino-acid is embedded as a kind of outsider for 64 DNA codes. In case of 80 DNA-24 generalized amino-acid code this imbedding is replaced with the imbedding of 2 amino-acids to 4 and 12 to $16 + 2$ structure.

This kind of imbedding would be regarded in the language of mathematician as a a discrete bundle structure which is also singular in the sense that the fiber above a given base point does not always have the same number of points. The 10×2 and 16×4 compositions suggest the interpretation as the imbedding of the space formed by 10 points of 2-D space-time to the space formed formed by 16 points of 4-D space-time. Analogous interpretation applies also in the case of the extended codes.

The interpretation conforms with the general idea that DNA represents a plan and involves intentionality and time dimension somehow. The amino-acids coded by several DNAs correspond to surfaces for several time values correspond to the same spatial point represented by amino-acid. The set of DNAs coding single amino-acid brings in mind the notion of "association sequence" defined as a disjoint union of space-like 3-surfaces with time-like separations and possible by the classical non-determinism of the Kähler action absolutely crucial for understanding consciousness in TGD framework [K10]. The number of DNAs coding the amino-acid would measure the degree of intentionality involved with it: each DNA associated with the amino-acid would symbolize one step in a plan. Some of alien amino-acids would be highly intentional: the degeneracies can be as high as 13 to be compared to the maximal degeneracy of 6 for our code!

Consider now in more detail this structure.

1. Exact A-G gauge symmetry implies that the pairs (XYA, XYG) form fibers and one can choose freely XYA or XYG to represent the amino-acid. In case of T-C symmetry symmetry breaking can select either XYT or XYC uniquely as a representative of the amino-acid.

2. For amino-acid coded by two DNAs only the identification of the amino-acid is unique apart from the possible gauge symmetry. For $n > 2$-plets the identification involves non-uniqueness.

3. The requirement that the imbedding of amino-acids to DNA space is universal allows to fix identification uniquely in case of $n > 2$-plets. It turns out that one can assume universal imbedding to make sense for both terrestrial and alien codes (if the replacements cys$\rightarrow$ scys and lys $\rightarrow$ plys possibly occurring for the small letter + special sign code are appropriately interpreted). This assumption fixes the imbedding highly uniquely and the only uncertainties relate to the T-C symmetry breaking. The possibility to choose the universal part of the code table to be the same for all codes, suggests that the proposed model catches something essential. It is also difficult to imagine that a randomly generated ASCII message could allow interpretation in terms of genetic codes having so high symmetry properties and common construction principles. **Table 18** summarizes the universal part of the genetic code resulting from the imbedding of the amino-acid space to DNA space. Also small letter code is included.

4.7 Summary

To sum up, both the terrestrial and hypothetic alien genetic codes can be constructed from the A-G and T-C symmetric product codes by assuming a breaking of both product– and T-C symmetries. Product structure and symmetries suggests strongly that genetic codes have evolved as a fusion of much simpler doublet and singlet codes. Hydrophilic-hydrophobic dichotomy is a good candidate for the dichotomy implied by the 2×10 product structure. The assumption that the breaking of the product symmetry induced by the "volume preserving flow" in DNA space tending to cluster amino-acids in the vertical direction of the code table is universal, and the hypothesis that the imbedding of the amino-acid space to the DNA space is universal, together fix the identification of the codes highly uniquely.

The small letter-special symbol code with 80 DNAs and 23 amino-acids is favored because it maximizes both the information content and the expressive power of the code. The degenerate

	A	G	T	C	T_S	
A	phe	ser	tyr	cys (scys)	.	A
						G
						T
	leu		stop	trp		C
G		pro	his	arg		A
						G
						T
			gln			C
T	ile	thr	asn			A
						G
						T
	met	thr	lys (plys)		&	C
C			asp			A
						G
						T
	val	ala	glu	gly	!	C

Table 18: A possible imbedding of the amino-acid space to the DNA space. The gauge choice XYA allowed by A-G gauge invariance of the last codon is made. The identification is same for both our code, capital letter code, and small letter plus special sign code. There is some uncertainty related to the T-C symmetry breaking.

code with 80 DNAs and 20 amino-acids is obtained from the 23-amino-acid code by assuming that the exotic DNAs coding for special signs !, & and period code for stopping sign. To my own opinion the OPpose option for the small letter code with 80 DNAs and 23 amino-acids is the most plausible alternative.

Acknowledgements

I want to express my gratitude for several persons. In particular, for Tapani Koivula for encouraging me to take seriously UFOs and ETs and for interesting suggestions (in particular, for stimulating the idea that tectonic energy could serve as a "food" of plasmoid like life forms), Martin Keitel for helping me to realize that crop circles are real and telling about the Chilbolton and Crabwood crop circles as well as for concrete help, and for Toni Siira for providing material related to UFOs and for interesting email exchanges. I want also to thank for Jukka Kinnunen for two champaign bottles (the first one for the capital letter code and the second one for the small letter code): I hope that I will sooner or later invent the means of establishing communications with aliens (the third bottle of champaign). I would also want to thank the ITs responsible for these charming Chilbolton and Crabwood crop formations. If I only knew their names! As an underground intellectual myself, I hope that amplitude modulated micro-waves or whatever from my computer could mediate my deep gratitude and warm greetings to this underground intelligentsia.

5 Genes, Memes, And Universal Language

In TGD framework the notion of magnetic body plays a key role in the understanding of bio-systems and communications between magnetic and material body could be based on memetic code words. Magnetic body is the fundamental intentional agent is in the same relation to the material body as memetic code to the genetic code and computer software to the hardware or the manual to the electronic instrument. Hence there is a strong temptation to believe that memetic codewords represented as field patterns of duration.1 seconds is associated with the communications between magnetic body and brain.

For instance, the fact that 10 Hz is basic hippocampal frequency suggests that declarative mem-

ories could be based on time mirror mechanism (see **Fig.** `http://tgdtheory.fi/appfigures/timemirror.jpg` or **Fig. ??** in the appendix of this book) with negative energy signal from the magnetic body of the geometric future reflected from the brain of the geometric past as a positive energy signal back to the magnetic body of the geometric future. Classical communications could utilize memetic code using pulse or frequency coding. These field patterns could also define shared mental images. For instance, sequences of the memes represented as field patterns could activate intronic memes, which in turn would activate genes.

5.1 Genes-Memes, Biology-Culture, Hardware-Software?

The reports of the Public Consortium about human genome in Nature, Feb 15, 2001 [I10] and of Celera Genomics in Science of Feb 16th, 2001, [I16] demonstrated that the amount of human genome differs relatively little from those of lower organisms: we have only about 30, 000 genes, little more than twice the number 13, 601 of genes for fruit fly. This paradoxical finding strongly supports the view that our genome is not solely responsible for what we are and that the intronic portion of DNA (only about 1 per cent codes of human DNA codes or amino-acid sequences), is not " junk DNA", but contains important biological information and expresses it non-chemically.

In TGD Universe introns would express memes as the classical field patterns associated with MEs ("topological light rays") responsible for the basic expressions of language understood in an extremely general sense. This language includes body language and even cellular signalling, and could quite well make possible (not necessarily conscious) interspecies communications based on the memes and genes expressed by both communicating species and forming a common portion of grammar and vocabulary. All eukaryotes (cells with nuclei), even bacteria, would possess part of the memes of this universal language. The memetic code word is predicted to consist of a sequence of 21 DNA triplets and carries 126 bits of information instead of 6 bits of genetic code. Of course, also genes could be expressed in terms of MEs and could define a lower level language possessed also by prokaryotes.

The actual role of DNA could be understood using a computer analogy. Memes represent the program modules written using the programming language defined by the memetic code and realized in terms of the field patterns associated with MEs. Genes represent the lower level programs coding for the necessary hardware. System builds only the hardware needed, that is cell expresses only a small fraction of the genome. For neurons this fraction is known to be highest. DNA engineering requires besides the addition of the new programs (memes, introns) also the insertion of the necessary hardware (new genes). Memes and corresponding genes should have very intimate relationship. In this conceptual framework the standard view is wrong since it identifies the build-up of a new hardware as the sole activity at the DNA level. This would be like identifying the addition of a net card to a computer as the fundamental activity related with computers.

5.2 Pulse And Frequency Representations Of The Genetic And Memetic Code Words

The most general form of p-adic length scale hypothesis implies that each p-adic prime $p \simeq 2^k$, k integer, defines a hierarchy of physically favored p-adic time scales given by $T_p(n) = p^{(n-1)/2}T_p \equiv T(n,k)$, $T_p = \sqrt{p}T_{CP_2}$, where T_{CP_2} is the so called CP_2 time scale about 10^4 Planck times. The most general assumption assigns to any prime $p \simeq 2^k$, k integer, a hierarchy of cognitive codes with codeword having a duration equal to n-ary p-adic time scale $T_p(n)$ such that the number of bits is factor k_1 of k.

Code words could be realized either as k_1 first harmonics of the fundamental frequency $f_p(n) = 1/T_p(n)$ or as temporal sequences of k_1 bits of duration $\tau = T_p(n)/k_1$ represented as pulses of maximal duration τ. These representations will be referred to as frequency and pulse representations respectively. EEG represents a good candidate for frequency representation.

1. Pulse representations, scalar wave pulses, and transformation of intentions to actions

Pulse representations could be realized in terms of scalar wave pulses predicted by TGD and claimed to exist already by Tesla [K3]. Scalar wave pulses can be visualized as capacitors moving with light velocity and carrying longitudinal essentially constant electric field. If charged particles of matter end up temporarily to the space-time sheets of the scalar wave pulse, they are

accelerated without dissipation and generate negative energy "acceleration radiation" rather than brehmstrahlung at harmonics of the frequency determined by the duration of the scalar wave pulse. Under obvious conditions on the duration of the scalar wave pulse the negative energy radiation can be amplified to positive energy radiation by time mirror mechanism.

Quite generally, the generation of scalar wave pulses seems to provide a basic mechanism generating negative energy radiation (phase conjugate radiation) and nerve pulses are probably accompanied by scalar wave pulses. The transformation of the p-adic counterpart of the space-time sheet of the scalar wave pulse to a real one, a kind of switch-on process, could provide a generic realization of intention besides a direct generation of the p-adic counterpart of the negative energy topological light ray. The hierarchy of magnetic bodies could use this process as a generic manner to realize cascades of intentions proceeding from magnetic body down to the level of DNA.

Denoting by $f(\omega)$ the Fourier transform of single pulse in the interval $T(n,k)$, one can write the Fourier transform of the pulse sequence as

$$F(\omega_n) = \sum_k \delta_k exp(ik\tau_b\omega_n) f(\omega_n) \quad , \quad \omega_n = \frac{n2\pi}{T(n,k)} \quad ,$$

where τ_b is the duration of the bit and δ_k is equal to 1 or 0 depending on whether k^{th} bit corresponds to a pulse or not. The duration of the pulse can be anything in the range $(0, \tau_b)$. If $f_b = 1/\tau_b$ corresponds to a frequency of some oscillation a resonant coupling occurs. Magnetic transition frequencies and the frequencies corresponding to the increments of zero point kinetic energies are especially interesting as far as the transformation of the pulse representation to a conscious experience or controlled action is considered. For instance, pulses could correspond to magnetic pulses used in the transcranial magnetic stimulation and known to induce altered states of consciousness [J12].

2. Genetic and memetic codes as cognitive codes associated with spoken and written language

Genetic and memetic codes are the most obvious candidates for the codes associated with spoken and written language. Genetic code would correspond to $k = 2^7 - 1 = 127$ and one must distinguish between 6-bit(64 DNA triplets) and almost-7-bit representations. These codewords can be realized dynamically as temporal field patterns. For genetic code primes $p \simeq 2^k$, $k = 6 \times n$ define candidates for the duration of the genetic code word if all factors of k are assumed to define a possible number of bits of the code word. The time scales come as powers of 8 so that they cover the entire range of biologically relevant time scales and genetic code could appear as fractally scaled versions unlike memetic code. What is interesting is that the possible durations of code word range down to about 11 CP_2 times. Therefore one cannot exclude the possibility that the biological realization of the genetic code is only a particular example of its realizations and that genetic code makes possible communications even between living and so called non-living matter.

5.2.1 Representations of the genetic code

$k = 2 \times 126 = 2 \times 6 \times 21 = 252$ allows the representation of both 126-bit memetic codeword, 6-bit genetic codeword, and 7-bit code word. For pulse representation corresponding to $k = 252 + 6n$ the genetic codon the duration of the code word and bit are $\tau = 2^{-3n} \times 50$ ms and $\tau_b = 2^{-3n} \times 8.3$ ms respectively. The realization using nerve pulse patterns certainly possible for $n \geq 0$, $n = 1$ would $\tau_b = 1.04$ ms which seems to be somewhat too short. Frequency representation would be realized using the 6 first harmonics of the fundamental frequency $f_1 = 2^{3n} \times 20$ Hz. In the following only 6-bit representations are discussed.

1. Representations of 6-bit code in the range of audible frequencies

20 Hz corresponds to the lowest end of audible frequencies ($20 - 2 \times 10^4$ Hz). Audible range allows 3 representations of the genetic code corresponding to the fundamental frequencies f_1 equal to 20 Hz, 160 Hz and 1280 Hz. 1 kHz frequency is between the frequency ranges associated with the latter two representations. Above (below) 1 kHz the wavelengths of incoming sound waves are shorter (longer) than head size so that the mechanisms determining the direction of the sound source are different above and below this frequency. Speech might correspond naturally to pulse representations whereas music could correspond to frequency representations. Also nerve pulse-EEG dichotomy could correspond to talking-singing dichotomy (left brain speaks and right brain sings).

2. EEG and nerve pulse representations of 6-bit code

1. Cortical EEG frequency range favors the realization using $f_1 = 2.5$ Hz: all harmonics in the range 2.5-15 Hz are important EEG resonance frequencies. For 20 Hz representation 120 Hz would represent the highest harmonic and it is questionable whether cortical EEG contains it with a sufficient intensity. Cerebellar EEG however allows much higher frequencies than cortical EEG.

2. The cyclotron frequency of Si^{++} ion is 21.4 Hz so that it could define a frequency representation of the genetic code in Earth's magnetic field, at least if its value is subject to a homeostatic control.

3. Ca^{++} has cyclotron frequency of 15 Hz in Earth's magnetic field with value $B_E = .5$ Gauss. A pulse representation of the genetic code with $\tau_B \simeq 66.7$ ms ($f_b = 15$ Hz) would excite harmonics of Ca_{++} cyclotron frequency and thus couple the representation to the Bose-Einstein condensate of Ca_{++} ions. Nerve pulses could realize this representation. Blackman [J6] has found that the harmonics of $f_b = 15$ Hz frequency have effects in living matter, and Ca^{++} waves are known to play an exceptional role in biology [J11] (for TGD based model see [K7]. Hence the Bose-Einstein condensate of Ca^{++} ions might provide a fundamental pulse representation of the genetic code.

4. The presence of fractally scaled-up versions of the Earth's magnetic field the sheets of the many-sheeted DNA would allow also scaled versions of Ca^{++} representation with durations $\tau = 8^{-n} \times .05$ ms of the code word. Depending on whether magnetic flux quanta are tubes or sheets the p-adic primes of space-time sheets of magnetic flux tubes comes as $k = 169 - 3n$ or $k = 169 - 6n$. For sheets one would have $k = (169 = 13^2, 163, 157, 151)$: these primes define important p-adically scale up electronic Compton scales relevant to DNA in the range 10 nm-5 μm and three of these primes define Gaussian Mersennes. Sheet option is favored by the explanation of the findings of Peter Gariaev about radio emission induced by irradiation of DNA by laser light [I13]. All these representations except $k = 169$ are realizable also using audible frequencies, which suggests a direct connection between the sheets of the many-sheeted DNA and the representations of the genetic code at audible frequencies.

5.2.2 Representations of the memetic code

For the memetic code one must distinguish between almost-127-bit representations and 126-bit representations. In both cases there is a very limited number of representations, which suggests that the emergence of memetic code might relate to the emergence of explosive cultural evolution. The first representations corresponds to the time scales $T(126)$ and $T(127)$: the latter defines the Compton time of electron. Next representations corresponds to the time scale of about .05 seconds and.1 seconds respectively.

1. 126-bit representations

1. 126-bit memetic code word can be represented using the same representation as for the genetic code, namely $k = 2 \times 126 = 252$ with $\tau = .05$ ms and $\tau_b = .4$ ms and $f = 20$ Hz and $f_b = 2520$ Hz. Audible frequencies could realize both representations and music experience might involve both frequency and pulse representation of 126-bit memetic code corresponding to the left brain (rhythm) and right brain (melody) aspects of music.

2. 126-bit representation using nerve pulses as such is not possible. The analog of the intronic representation of the memetic code using sequences of 21 DNA triplets could be however possible. $k = 252$ allows 21-bit representation for which bit is replaced by 6-bit with duration $\tau_6 = 50/21 = 2.38$ ms, which corresponds to a typical duration of nerve pulse. Each of the 21 nerve pulses should generate a genetic codon of duration $\tau = 50/64 = .78$ ms presumably communicated to the neuronal nucleus and/or vice versa. 6-bit fine structure could be perhaps expressed at the microtubular level as has been originally proposed by Koruga [I11], [J8]. One would have $\tau_b/\tau_6 = 21/64$, the ratio of the number of amino-adics (stopping sign counted as effective amino-acid) to the number of DNA codewords. τ_b/τ_6

represents a reduction of information: this loss of information is not due to the degeneracy of the code but due to the fact that only one third of the total duration of the bit of 21-bit code is used to represent information.

One could imagine that the 6 bits represented as negative energy topological light rays activate the corresponding DNA triplet by coupling 3 switches on so that supra-current can flow through it. Sequence of 21 pulses would switch a unique memetic codon and sequences of these pulses in turn would switch on meme forming by definition a closed supra current circuit with return current flowing along conjugate strand. Switches could correspond to join along boundaries bonds connecting atomic space-time sheet to some larger space-time sheet, where the supra-current can flow. Each pair of bits would switch on one nucleotide of the triplet. This would occur in correct order if the three nucleotides are ordered by pulse length associated with bit (which can indeed vary) and 4 different bit pairs switch on A, T, C, and G. Bit could be represented by the polarization direction associated with the negative energy topological light ray.

The generation of nerve pulses using 64 bit sequence to code additional information at microtubular level could be based on frequency representation. Pairs of em MEs with opposite polarizations could represent a bit pair corresponding to a single nucleotide. These waves would induce microtubular excitation representing the DNA triplet.

2. 127-bit representations

For 127-bit representation the duration of the memetic codeword would be $T(2, M_{127}) = .1$ seconds. This time scale might be identified as the minimal duration of cortical mental images, and the so called features introduced by Walter Freeman [E1] could define a pulse representation of memetic code words of almost-127 bits. $\tau_b = .8$ ms is definitely too short a time scale to be realized by the neuronal dynamics alone. Frequency representation is realized utilizing 127 first harmonics of $f_1 = 10$ Hz, which defines the average frequency of alpha band and is fundamental hippocampal frequency. $f_b = 1270$ Hz could define the frequency responsible for synchronous neuronal firing known to be about 1 kHz. Note that the code word containing only f_1 would not generate any conscious experience (10 Hz is not audible frequency) so that the highest bit is not quite fully represented.

1. One can imagine at least two electromagnetic realizations (for the spectrum of magnetic transition frequencies see [K13]).
   i) Living matter contains both Co and Fe ions and the harmonics of Co_{++} and Fe_{++} cyclotron frequencies are 10 Hz for the nominal value $B_E = .5$ of the Earth's magnetic field. Thus pulse both pulse representations and frequency representations of memetic code coupling to these magnetic transitions are in principle possible.
   ii) For $B = 127/90 \times B_E$, $B_E = .5$ Gauss, both the third harmonic of proton cyclotron frequency and electron's spin-flip-cyclotron transition frequencies are 1270 Hz so that the bits of the memetic codon would couple to the Bose-Einstein condensate of the Cooper pairs of electrons and protons in the pulse representation.

2. Also classical Z^0 fields make possible realizations of the memetic code (for the spectrum of Z^0 magnetic transition frequencies see [K13]).

   (a) Z^0 cyclotron frequencies of nuclei are proportional to $(A - Z)/A$ and around 10 Hz if one assumes the earlier hypothesis that Earth's Z^0 magnetic field corresponds to the space-time sheet $k = 173$ [K12]. The value $(A - Z)/A = 1/2$ characterizing surprisingly many biologically important ions (C, N, O, S and Si [K19]) is ideal in this respect. Thus it would seem that Z^0 cyclotron transitions might provide a rich repertoire of frequency representations of the memetic code. Note that neutron could define a cyclotron representation of the genetic code with 20 Hz fundamental frequency.

   (b) The temporal field patterns associated Z^0 topological light rays provide one possible pulse representation of the memetic code [K13].

   (c) Temporal sequences for the changes Z^0 magnetization directions for a block of cognitive antineutrinos at cell membrane space-time sheet provide a conscious pulse representation of the memetic codeword [K12]. Conscious experience would result, when the Z^0

magnetization directions flip back to the direction of external Z^0 magnetic field in spin flipping cyclotron transition. $M_{127} = 2^{127} - 1$ different conscious experiences results since nothing happens if all cognitive antineutrinos are in the direction of the external Z^0 magnetic field. Z^0 magnetization direction could be altered by the Z^0 magnetic pulse associated with the Z^0 ME inducing cell membrane oscillations of nerve pulse pattern.

5.3 Mapping Of The Memetic Code To Microtubular Code

The importance of microtubuli for long term memory is evident [J13, J8]. Microtubule decomposes into a sequence of cylinders containing 13×13 tubulins such that the helical twist is 2π along each of the 13 helical strands consisting of 13 tubulins. Therefore the code with $k = 13^2 = 169$ bits with bit realized as a tubulin conformation is a natural microtubular cognitive code [K11].

$k = 169$ defines the p-adic time scale associated with the Earth's magnetic field and next to the p-adic length scales $k = 151, 157, 163, 167$ associated with DNA so that microtubular level would naturally correspond to the level next to DNA in evolution, and have therefore some sort of self-reflective character. Microtubular representation of long term memories would certainly be consistent with this self-reflective character. This suggests that 169-bit microtubular code words represent the log file of neuron as a temporal list of activated 126-bit memetic code words with remaining bits representing parity bits making possible error correction at both microtubular and DNA level.

In the following some arguments for why microtubular code words should represent memetic code words are developed, and a mechanism for how to achieve this is proposed. Needless to say, this is only one possible scenario and it is easy to imagine variants of this scenario.

5.3.1 Microtubuli and long term memory

In spin glass phase tubulin conformations are spatially uncorrelated but temporally stable (the excellent articles of Dimitri Nanopoulos [J13, J8] provides a model for microtubule as spin glass). Therefore microtubuli in spin glass phase are ideal for the representation of memories coded to bits represented by tubulin conformations [K11], [J8]. The two tubulin conformations have different electric dipole moments and conscious bits would result as "spin-flips" when the microtubule is in a strong longitudinal electric field forcing tubulins to the same conformational state. Essentially ferro-electric polarization is in question. The fundamental quale would be the change of the tubulin conformation. The patterns for the changes of tubulin conformations would generate mental images, and could also give rise to conscious memories by sharing of mental images. They could also give rise to signals communicated classically to the geometric future where they could induce reverse transition generating copies of the microtubular code words.

Time mirror mechanism for the realization of intentions proceeding as a process initiated from the magnetic body allows to consider two possible options for the role of the microtubuli. The options are not of course mutually exclusive.

1. Microtubuli as log files and communication lines

For this option intentional action does not involve microtubuli and they would be specialized to represent memories and serve as communication lines. The intentional action from higher than microtubular level would affect intronic DNAs directly, and patterns of tubulin conformations would provide a log file listing the memetic codons activated during the history of the neuron. Since the hierarchy of the magnetic bodies must have been there for all the time, one an indeed argue that neuronal microtubules have emerged later and do not participate in the intentional actions at intermediate level. Intentional action and the memory about it would be decoupled from each other completely.

Microtubuli allow besides ferro-electric and spin glass phase also phase which is optimal for signal transfer. The proposed realization of memetic code words as sequences of 21 nerve pulses with each pulse accompanied by genetic codon would suggest that microtubuli also mediate propagation of memetic codon, which can represent the desire to activate corresponding memetic codon in the post-synaptic neuron. Here the error correcting code $K_2(13, 64, 5)$ originally proposed by Koruga [I11], [J8] could be involved.

2. Why microtubuli cannot serve as intentional agents?

One must also consider the option for which microtubuli would represent the last step before DNA level in the hierarchy of desires propagating downwards in the self hierarchy. One can however represent heavy criticism against this alternative.

1. One can wonder whether the microtubular memes are generated intentionally or in a random manner in a phase transition leading to spin glass phase with basically un-predictable meme sequence. In the latter case, intentional action would be reduced to a selection to activate or not to activate the existing memes.

2. In this case it would be possible to have long term memories about events that never occurred, which seems strange. Random generation of the memes would also be in conflict with the notion that there are at least 42 parity bits making possible error correction. Thus it would seem that microtubular codewords can be activated only from the DNA level. In the case that microtubuli act as signal pathways this would indeed be the case.

3. Not all memetic code word sequences representable at microtubular level need to have counterpart at DNA level. This would lead to a situations in which meme could not be expressed at all.

These arguments favor the view that microtubuli are passive historians making possible self-reflection by providing a log file about activated memetic codons and possibly serve also as communication lines allowing the propagation of memetic codons between neuronal nuclei as sequences of 21 nerve pulses accompanied by genetic codon each. Only this option will be discussed in the sequel.

3. How to generate and read microtubular code words?

The coding of the intronic memetic code word to a microtubular code word would involve switching-on mechanism in spin glass phase of the microtubule for which initial state consists of "0": s. The tubulins corresponding to bit "1" would make transition to the conformation representing bit "1". The activation of intronic meme should automatically generate the positive energy photons at frequency corresponding to the energy difference between two conformations of tubulin. Intronic memes should have kind of hardwired connection to a fixed ordered sequence of microtubular code words. Note that in TGD framework there is no need to static microtubular memory since memories can be communicated from geometric past. Therefore memory capacity would be unlimited in this sense.

The conscious reading of the microtubular code word using strong enough longitudinal electric field would generate positive energy photons, which could be communicated to the geometric future and generate declarative memory mental images. Also a direct sharing of mental images yielding episodal memories is possible.

5.3.2 Representation of the memetic code words as microtubular code words

The challenge is to understand how 126-bit genetic code word or (possibly 127-bit codeword) is mapped to 169-bit microtubular codeword. There are several hints how this mapping could be realized.

1. Number theoretical decomposition of the microtubular code word to memetic code word and 43 parity bits

It is possible to represent microtubular 169-bit codewords as $13 \times 13 = 169$ square lattice of bits.

1. One can write the number of microtubular bits as

$$169 = 126 + 43 = 3 \times 42 + 42 + 1 \ ,$$

and there is a temptation to assume that the first 126 bits correspond to the memetic codon and 42+1 bits represent parity bits making possible error detection. From the geometric

representation it is clear that the bit in the middle of the 13×13 square is an excellent candidate for "1" in $42 + 1$.

2. 126-bit memetic code word allows a natural identification of the parity bits. The $126 = 3 \times 42$ decomposition allows also 42-bit code word, whose bits are obtained by decomposing 126-bit code word to a sequence of 42 3-bits, and defining each bit B of 42-bit as some Boolean function $B = f(b_1, b_2, b_3)$, say as a product $B = b_1 b_2 b_3$ of the bits of the corresponding 3-bit by representing bits as 1 and -1: the result is -1 (bit "0")if the number of bits "0" is odd and 1 otherwise. The comparison of codewords with a reference codeword assumed to be the correct one would allow to locate errors leading to internal inconsistency of the code word. The comparison of 42-bit codeword with the original one would allow to locate single bit changes of the memetic code word with a resolution of 3-bits.

3. A possible interpretation of the 43^{th} bit results from the requirement that each memetic code word gives rise to a conscious experience. This is guaranteed if the bit in the middle of the square is always "1" so that in the phase transition it changes the direction always and conscious experience results even when all the remaining bits are "0": s: the interpretation in this case would be as a mental image representing "nothing happened". An alternative possibility is that this bit represents parity bit. For instance, the product of all memetic bits or of all 42 parity bits. This bit could also be "1" only if there are no erratic bits. This bit could also represent 127^{th} bit of the 127-bit memetic code word.

2. How intronic DNA could represent parity bits?

If microtubule represents passively the information communicated to it, intronic memes should be accompanied by 42-bit parity code words. The minimal portion of DNA helix containing an integer number of DNA triplets consists of 10 triplets and corresponds to a length of $L(151) = 10$ nm (cell membrane thickness). 20 triplets would correspond to 2 full 2π twists. This encourages to consider the possibility that memetic codons correspond to 3 2π twists (length of 30 nm) along DNA so that memetic code words are followed by by a sequence of 9 DNA triplets serving some control function. Since 10 nm corresponds to electron Compton scale $L_e(151)$, the idea that dark electrons with this Compton length are relevant.

These 9 DNA triplets represent 54 bits. 7 of these codons could represent the 42 parity bits of the memetic codon preceding it. The remaining two genetic codons (12 bits) should represent further control information. The factorization $Z_{126} = Z_2 \times Z_3^{a)} \times Z_3^{b)} \times Z_7$ allows to identify two coset groups $Z_{126}/Z_{21} = Z_6$ as $Z_6^{a)} = Z_2 \times Z_3^{a)}$ and $Z_6^{b)} = Z_2 \times Z_3^{b)}$ corresponding to the identifications $Z_{21}^{a)} = Z_3^{a)} \times Z_7$ and $Z_{21}^{b)} = Z_3^{b)} \times Z_7$. This would mean the possibility to define two non-equivalent 6-bit parity code words as products $B = b_1 \times ....b_{21}$ of 21 memetic bits corresponding to the two sub-groups Z_{21}. The two DNA triplets could represent these parity 6-bits. Needless to say, these prediction are very strong and immediately testable by studying the intronic DNA.

3. The helical structure of the microtubule and the representation of the memetic code word and of parity bits

The tubulins representing parity bits should differ physically from those representing memetic bits. Time mirror mechanism suggests that the energy difference between two tubulin conformations differs considerably from that for the memetic tubulins. A weaker symmetry breaking induced by the helical electric field should order the bits of helical genetic code words and also order the 7 genetic code words to a sequence of vertical pairs running from left to right. Genetic code words should correspond to connected regions, most naturally helical stripes. Also the memetic code word and parity bits should correspond to a connected regions. There should be a clear signature telling where the microtubular code word ends.

The breaking of the rotational and translational symmetries is necessary and the helical structure of microtubuli could induce it. Denote by a the vertical distance between tubulins and by $\Delta\phi = 2\pi/13$ the angular distance between two tubulins along horizontal circle.

1. Genetic codons would be arranged to the helical stripes rotating making full 2π around the vertical section section of the microtubule defined by 13×13 tubulins. The 10 helices span a helical region spanning the angle range $\Delta\phi = 10 \times 2\pi/13$. Each helix would represent

2 memetic codons whereas the 13^{th} tubulins at $z/a = 13$ would represent parity bit. This would give 20 genetic codons. The 21^{th} genetic codon would correspond to the 6 lowest tubulins of the 11^{th} helical stripe. Therefore the pairs of 2 sub-sequent genetic code words would correspond to helices beginning at

$$(\tfrac{z}{a} = 1, \phi = k \times \Delta\phi) \ , \quad (\tfrac{z}{a} = 7, \phi = k \times \Delta\phi + \pi) \ ,$$

$$\Delta\phi = \tfrac{2\pi}{13} \ , \qquad k \in \{1, 2, ..5\} \cup \{9, 10, ..13\} \ .$$

The 21^{th} code word would correspond to $z = na$, $n = 1, ..., 6$, $k = 11$. The region representing parity bits would be connected and consist of vertical part and the 13^{th} horizontal stripe. It turns out that this guess is probably quite not correct: it seems that the first pair of helices representing genetic codons starts at $(z/a = 2, k = 2)$ so that the horizontal line at $z/a = 13$ contains 12 tubulins (2 genetic code words).

2. Perhaps the simplest mechanism guaranteeing the desired symmetry breaking is based on a static helical electric field, which is non-rotational and thus representable as a gradient of a scalar potential V. Electric field could be constant in the memetic section of the microtubule since tubulins are charged and Coulombic interaction energy would grow linearly with the azimutal angle ϕ and longitudinal coordinate z along the helical strands. In this case the potential would be of form

$$V(z, \phi) = k_1 \times \frac{z}{a} + k_2 \times \frac{\phi}{2\pi}$$

in the region populated by the genetic code words.

The value of the potential should fall rapidly down at $z = 13 \times a$ defining the upper edge of the microtubular codon and the vertical stripes defining the parity bits. The strong gradient of the electric field in the parity bit region could give rise to a strong dipole interaction and change the energy difference between microtubular conformations dramatically both at $z/a = 13$ and inside the

   (a) lower helical stripe S_{low} having having height of 6 units and width of 2 units starting at $(z/a = 1, \phi = k \times \Delta\phi)$, $k = 12, 13$ and

   (b) upper helical stripe S_{up} having height of 6 units and width of 3 units and starting at $(z/a = 7, \phi = k \times \Delta\phi) + \pi)$, $k = 11, 12, 13$.

This behavior of electric field dictated by its non-rotational character would differentiate between memetic and parity bits.

3. Genetic codons are ordered to a sequence of pairs of codons if the value of the potential increases with constant steps along the helical stripe, and also in the transition from the top of k^{th} helical stripe to the bottom of $(k+1)^{th}$ helical stripe. This gives

$$k_1 = -\frac{k_2}{13} \ ,$$

which fixes the potential apart from overall scaling:

$$V(z, \phi) = V_0 \times (\tfrac{z}{a} - 13 \times \tfrac{\phi}{2\pi}) \ ,$$

giving

$$V(z = n \times a, \ \ \phi = s \times \tfrac{2\pi}{13}) = V_0 \times (n - 13 \times s) \ .$$

The length of a single tubulin in the vertical direction is $a \simeq 8$ nm and the outer radius of microtubule is $R \simeq 12.5$ nm [J13]. This gives $\Delta z/R\Delta\phi \simeq 1.32$ whereas $\Delta V_z/\Delta V_\phi = -1$ rather than $-1/1.32$ required by the orthogonality with respect to the helical strands. Thus the electric field can be regarded as a superposition of a field orthogonal to helical strands with a weaker field parallel to the z-axis.

4. The requirement that the 5+5+1 (5+5) microtubular codons are represented at the lower (upper) half of codon can be satisfied if the the value of the potential becomes in some sense too large at the point $(z/a = 7, \phi/2\pi = 11/13)$ and that this forces V to rapidly raise up from the minimum value

$$V_{min} = -137V_0 \ .$$

The maximum of the potential at $(n = 1, s = 1)$ given by

$$V_{max} = -12V_0 \ .$$

$(n = 1, s = 1)$ bit is non-memetic bit and shifted by one unit upwards from the horizontal $n = 13$ row of 13 bits: this distinguishes it as a natural candidate for the 43^{th} bit. With this identification the first helical stripe containing two genetic codons would start at $(n = 2, s = 2)$ and $n = 13$ horizontal line would contain only 12 points corresponding to two genetic codons. Vertical stripes of parity bits would contain 5 genetic codons represented as helical stripes. Thus all genetic codons would correspond to helical or horizontal stripes. Horizontal and helical genetic codons should differ somehow but the origin of difference remains unclear.

5.3.3 Symmetries of the genetic code and the error detection

The ratio of the number of the number of parity bits to memetic bits is $R = 42/126 = 1/3$ and almost equals to the ratio 21/64 of the number of genetic code words to the number amino-acids. If parity bits represent directly the "memetic amino-acids" the action of all memetic code words with given 42 parity bits is identical and the comparison of parity bits would be enough to check that everything is in good shape at the level of the memetic expression. Obviously, this option is trivial.

If the memetic code is directly induced by the genetic code, the ratio R would be $R = log_2(21)/6 \simeq .73$. For both options the number of the "memetic amino-acids" would be considerably smaller than $(2^{126} - 1)/63$ suggested by the direct generalization from the case of the genetic code, and corresponding to $R \simeq 20/21$.

If the memetic code is induced by the genetic code which is same for both chemical and field pattern expressions, intronic memetic codons possess the A-G symmetry and almost exact T-C symmetry of the genetic code in the sense that memetic code words related by these symmetries would activate the same genes. This would allow to improve further the error detection.

1. A-G and almost T-C symmetries imply that the last 2 bits of 6-bit defining genetic code word in good approximation reduce to single bit, say the last bit. Therefore the second parity bits could be taken to be the product b_4b_5. This improves considerably the effectiveness of error detection at the level of expression.

2. Also the first bit of the genetic code word is less significant since the genetic code table contains many 4-sub columns (codons of form XYZ, with Y fixed) for which all 4 amino-acids are same and quite a many 2-sub-columns (codons of form XYZ, X= A, G or X= T, C) are identical. Therefore a relatively safe choice is to take the first parity bit for the genetic codon to be the product b_2b_3. This choice would localize with a reasonable reliability the errors with 2-bit accuracy at the level of expression.

5.3.4 Could error correction mechanism be used to detect mutations of memes?

The first question to be answered concerns the purpose of the error correction process. Is error correction used

1. to remove the internal inconsistencies of the memetic code words at intronic level due to mutations or,

2. to stabilize long term memories represented at microtubular level?

Only the first option would be needed if the refreshment of the memetic code words by activated intronic DNA occurs all the subjective time in the geometric past. Both types of error correction could rely on the following mechanism.

The comparison of the microtubular code word with that of the geometric past by time mirror mechanism, assumed to yield resonant interaction if the compared code words are identical, could reveal the emergence of those mutations at DNA level which alter some parity bits and memetic bits in such a manner that the parity bit pattern ceases to be consistent with the memetic bit pattern. There would be simply no internally consistent microtubular code words in the geometric past identical with this pattern. Large number of temporal copies of the codeword in geometric past would make the mechanism reliable.

The internally inconsistent change of the microtubular codeword and even a change in general could be detected by this mechanism. In absence of resonance a process trying to fix up the meme at DNA level by trial and error could be initiated and could yield an internally consistent or even original meme. Note that this comparison mechanism would be conscious and even the correction mechanism might be conscious intentional process.

To sum up, the model for the mapping of memetic code to microtubular code is dictated by the general ideas about realization of intentions and p-adic cognitive codes. When combined with general number theoretical arguments and physical considerations the model becomes highly unique. The prediction for the intronic representation of the memetic codon is readily testable, and also the prediction for the microtubular electric field is in principle testable.

5.4 Genes, Memes, And Language

The idea about intronic memes as computer programs running in the hardware coded by genes together with quantitative facts lead to a view about how memes and genes could relate to language. It must be emphasized that the idea about correspondence between genes and language is not new. The article [I12] gives a nice summary about various findings supporting the view that language and DNA are closely related.

5.4.1 Zipf's law

One of the basic resemblances between genes and language is Zipf's law relating the rank r of the event with its frequency f of occurrence [J9]. The rank of a given event is defined by ordering the events in sequence using as a criterion the frequency of occurrence: the event occurring most often has rank 1, etc.. Zipf's law states $f = C/r^a$, where a is a constant near one, and C is a constant depending on the size of the sample. In linguistics the event could correspond to the occurrence of a given word or character, and in genetics to the occurrence of a given DNA triplet.

Zipf's law holds also for many systems regarded as non-living, such as masses of stars in a constellation of stars. Interestingly, the sum of the frequencies for an infinite number of events is proportional to Riemann Zeta function $\zeta(a) = \sum_n 1/n^a$ at point $a \simeq 1$, which can be regarded as a product of thermal partition functions associated with primes regarded as bosons: $\zeta(a) = \prod_p 1/(1-p^a)$, $log(p)$ plays here the role of energy and a to the role of inverse temperature. Interestingly, the line $a = 1$ defines a critical line such that $\zeta(a)$ converges for $Re(a) > 1$. $a = 1$ corresponds to the temperature at which Bose-Einstein condensation occurs and the system becomes macroscopic quantum system. Thus Zipf's law might relate to p-adicity and macro-temporal quantum coherence in some deep manner.

There is also evidence in favor of a typological relationship between the words of human speech and DNA "words" (see the references in [I12]). The word formation appears to obey laws similar to those of formal genetics so that one can speak of dominant and recessive features, mutations, etc.. In [I12] graphical representations character sequences defined by written texts and nucleotide (A, T, C, G) sequences are studied (amino-acid sequences of long proteins would be more appropriate for comparison with character sequences). The fractal dimensions of the resulting planar fractals are in the range $0 < D < 1$. A random sequence of characters corresponds to $D = 0$ and repetition of single character to $D = 1$, whereas written texts correspond to $D = .7 - .8$.

5.4.2 Could the phonemes of language be expressions of DNA triplets?

The previous construction of representations of genetic and memetic codes, in particular the facts that the frequency $f \sim 10$ Hz represents the fundamental frequency associated with speech organs and that 20 Hz frequency represents the lower limit for audible frequencies and the rate of mRNA-amino-acid translation, suggest that memetic code might be involved with speech and speed recognition. Since the duration of a phoneme in speech is $\sim .05 - .1$ seconds, one must seriously consider the possibility that phonemes might somehow relate to genetic or memetic codewords.

The number of phonemes of native origin in Finnish language the same as the number of characters and equals to 21, the number of amino-acids plus stopping codon counted as an effective amino-acid. In English language the number of phonemes it is not so easy to estimate the number of phonemes since written language differs so much from spoken but the number is not too far from 21. In Russian the number of different characters is 32, half of the number of DNA codons. Certainly one cannot expect precise correspondence between the numbers of phonemes/letters and amino-acids and it might well be that DNA→phoneme coding differs from the DNA→amino-acid coding.

These findings together with the observation force to consider the possibility that phonemes might play a role similar to that of amino-acids, and be interpreted as resulting in the translation of DNA triplets to field patterns to neural activity to the motor activity of speech organs. Words in turn would represent analogs of protein sequences coded by sequences of DNA triplets. The fact that memetic code represents statements about statements would suggest that memes represented by introns correspond to higher level structures of language, kind of main computer program producing meaningful expressions of language while running. Words represented by genes could be the hardware of language or lower level language and DNA triplets would represent the basic material of the hardware expressed as phonemes. This would of course require expression of genes and memes in terms of field patterns inducing the generation of nerve pulse patterns.

Consonants carry more information than vowels (early form or written language used symbols only for consonants). This picture is consistent with the fact that the property of being consonant is determined by the very brief time interval in the beginning of the pulse sequence defining the phoneme so that the occurrence of high frequencies make it consonant. There are 8 (7) vowels in Finnish (English) language, which suggests that the first 3 bits of the generic codeword in the frequency representation and representing lowest frequencies are responsible for the basic vowels and the remaining 3 bits corresponding to the higher frequencies add to the beginning of the phoneme something making it consonant.

The frequencies for the occurrence of various phonemes should be the same as for the occurrence of the DNA triplets coding them. The simplest working hypothesis that the DNA-phoneme code is same as DNA-amino-acid code does not work in case of say Russian language (32 characters) but deserves a testing in the case of Finnish language. Of course, one could consider the possibility of context dependent coding of phonemes: also ordinary genetic code can be context dependent in some situations (the same codon codes for stopping sign or amino-acid depending on context). The prediction would be that the frequencies for the occurrence of phonemes would equal to those of amino-acids.

It is not possible to assign a definite duration to the intronic representation of the memetic codeword. Speech is indeed extremely flexible: the rate of speech can vary, vowels can be represented multiply, etc.. This variation would be made possible by flexibility at the DNA level or at level higher than this (magnetic body would control introns). If speech represents genetic code, 50 ms would however be a good guess for the minimal duration of phoneme and 20 Hz for the lower end of consciously experienced frequencies (as it indeed is). Rather remarkably, the rate for the translation of mRNA triplets to amino-acids is 20 nucleotides per second so that one codon corresponds to the duration of the genetic codon in $k = 252$ representation of the genetic code.

On the other hand, the memetic code words generated while listening speech or reading text could correspond to the mental images representing the meaning of a word, and could quite well have a duration of.1 seconds whereas the minimal duration of phoneme would be about 50 milliseconds. If time mirror mechanism is involved time is not problem: the understanding of speech would occur at the level of DNA in the geometric past. Same applies also to the production of speech: the desire to produce an expression of language would generate DNA activity in the geometric past.

The overall conclusion is that memes and genes of DNA would define language independent universal grammar introduced by [J14] [J14]. Languages would differ from each other only in the different association of phonemes to the DNA triplets. Also languages based on electromagnetic realization and not directly conscious to us are possible and even probable.

5.4.3 How memes could control genes in the production of language?

Language is not expected to be the only function controlled by memes. Memes could be responsible for a high level control of genes also in their ordinary activities. Memes would represent a higher level control structure utilizing the existing program like structures defined by genes activating genes. In case of language the fundamental expression would be in terms of field patterns inducing nerve pulse patterns inducing in turn speech motor activities.

Computer metaphor suggests that memetic programs correspond to sequences of memetic codewords with each memetic codeword defining a main program as a sequence of sub-program calls. Production of speech would mean activating these programs from magnetic body by time mirror mechanism. Genetic program calls would be realized as addresses pointing to genes defining the subprograms: 2^{126} different addresses is much more than needed to point to any desired gene. Also genes could use this addressing to point to other genes. The pointed gene could generate field pattern giving rise to a word or sequence of words or it could point to another gene. In this manner meme could generate entire sequence of words by inducing a sequence of subprogram calls (computer language LISP suggests a possible model for what occurs). Genes have a hierarchical structure, and one expects that higher levels of this hierarchy could fix the parameters determining the rate of speech, durations of phonemes, accents, emotional content, etc....

Pointing could be achieved by electromagnetic field patterns utilizing time mirror mechanism so that address would be expressed by negative energy topological light rays (phase conjugate laser waves). Pointing could be based on avoidance of starvation. Negative energy photons at the frequency serving as the name of the pointed system would induce the dropping of charged particles to a larger space-time sheet.

In many-sheeted space-time particles topologically condense at all space-time sheets having projection to given region of space-time so that this option makes sense only near the boundaries of space-time sheet of a given system. Also p-adic phase transition increasing the size of the space-time sheet could take place and the liberated energy would correspond to the reduction of zero point kinetic energy. Particles could be transferred from a portion of magnetic flux tube portion to another one with different value of magnetic field and possibly also of Planck constant h_{eff} so that cyclotron energy would be liberated.

The pointed system would lose energy and would generate negative energy topological light rays at characteristic frequencies to compensate the lost energy. If the frequency is higher, the pointed system would point to a lower level in the hierarchy of sub-programs. One cannot of course exclude the possibility that pointed gene points back to the pointing gene or meme. This could give rise to recursive self-referential program structures. At the lowest level of hierarchy there would be gene which act as a population inverted many-sheeted laser generating a cascade of positive energy photons at various frequencies and in this manner yielding the desired response.

5.5 Does Memetic Code Make Possible Communications Between Different Species?

The observations of Cleve Backster [J5] give scientific support for the view that plants react to human emotions. The TGD based interpretation would be as sharing as mental images. One can however ask whether sharing of mental images might make possible the development of more evolved communications involving signalling by codes. For instance, people in hypnosis report that they can share the experiences of plants and shamans claim that they can experience what it is to be an animal and communicate with plants. According to what the shamans of the South America tell, their refined medical knowledge is told to them by plants [J10]. In principle common intronic memes could make possible this kind of communications even between different species: these communications would not of course be conscious to us under normal circumstances.

5.5.1 Interspecies communications by sharing of mental images

Sharing of mental images does not require neither target nor receiver to be able to communicate symbolically. Therefore the target and receiver could be any living system: animal, plant, even bacterium. In TGD Universe one cannot exclude even "non-living" systems as targets and even sharers of mental images. The remote viewing of non-living targets is indeed possible and in this case either mental images of target or some system perceiving target are shared. Since emotions represent holistic summaries about contents of consciousness, they are good candidates for universal mental images and the sharing of emotions could occur even between different species.

Support for the extreme generality of the sharing of the mental images as a basic mechanism of remote viewing comes the fascinating experimental discoveries made by Cleve Backster [J3, J5]. These findings have led Backster to introduce the notion of primary perception, which seems to have a natural identification as sharing of mental images.

1. Plants, eggs, and even bacteria are able to have primary perceptions. Backster tells in the interview that even yoghurt got wild when he took a chicken out of refrigerator and began pulling off strips of meat. Plants respond electrically to strong negative emotions and to the violence or death suffered by other living organisms. That primary perception correlates with the strength of emotions conforms with the view that entropy gradients with respect to subjective time, which are indeed identifiable as emotions, measure the strength of perception.

2. Distance does not seem to matter much. Sperm separated by a large distance from its donor reacted when the donor inhaled amyl nitrate. White cells were found to remotely react to the emotions of their donors. Same was found to apply to plants and their owners.

3. Plants and even bacteria seem to have a defence mechanism resembling shock. If bacteria share the mental images of suffering organisms by receiving negative energy MEs sent by them, the shock could be interpreted as resulting from the depletion of positive energy resources (all excited states of population inverted many-sheeted lasers decay to the ground state) or be a mechanism preventing this depletion.

An interesting question is whether humans have lost this ability or is this reaction usually unconscious at our level of self hierarchy and whether human skin could exhibit galvanic skin response to say death of other life-forms.

5.5.2 Interspecies classical communications using common memes and genes

The assumption that even classical communications between different species are possible is much stronger hypothesis than the assumption that mere sharing of mental images occurs. The general model for interspecies communications using symbolic representations for the communicated information rather than mere sharing of mental images relies on the pan-psychic view about life and consciousness ("Everything is consciousness and consciousness can be only lost"). Also the hypothesis that speech and written language are only one particular realization of language involving at fundamental level memes controlling genes expressing themselves in one of the many possible manners, supports this view.

1. General model

1. If the intronic portion of the DNA corresponds to memone defining basic linguistic repertoire expressed using genes defining the vocabulary. Even plants could possess language possibly realized using MEs ("topological light rays"). Introns could control genes in terms of electromagnetic MEs and genes could could express themselves by MEs. Genes would define the words of the language. One could say that interspecies communications reduce actually to intercellular communications between cells of different species.

2. Successful communication requires that the languages of communicators have common portion in their vocabularies. Cells possessing common expressed memes defining basic expressions and genes defining the vocabulary could communicate with each other. Meme level

communications would be possible between any two eukariotes (organisms having cell nucleus): bacteria, plants, fungi, animals. Also communications between plants and humans would be possible.

3. Communications could occur by sharing of DNA level mental images induced by activated memes: negative energy MEs would make possible the needed quantum entanglement. Communications could occur also classically. The simplest form of classical communication would involve the coding of the meme to the field pattern of em or Z^0 ME to a name of meme using memetic code in turn activating the same meme in the receiving organism. Memes could express themselves by controlling genes in turn generating various types of expressions for the meme This could in principle allow even translation of memes to speech in altered states of consciousness in which back projection to ears would transform the internal speech to sounds.

4. The general model for sensory representations and motor actions based on time mirror mechanism involves both quantum and classical communications. There is a great temptation to assume that the same model applies also in the case of inter-species communications. Both the psychokinetic aspect corresponding to the remote motor action aspect, and the remote sensing aspect corresponding to the remote sensory perception would be present and would be essentially geometric time reversals of each other. Psychokinetic aspect of course represents more than mere communications.

5. Unless the code is universal so that the classical signal induces directly a standard mental image, most plausibly emotion, irrespective of species, a code assigning symbols to mental images must be established by sharing of mental images. There are good justifications to argue that the establishment of a code requires sufficiently intelligent communicators. On the other hand, even if the signals from say plant do not have any "meaning" for a human receiver, they could generate effects analogous to synesthesia.

6. Organisms living in symbiosis (say some insects and flowers) could have common memes and genes allowing them to communicate. Common memes and genes would be the analogs of common genes causing the coloring of the flower and insect to be similar. This kind of communication could also explain the refined medicine of South American shamans. Shamans tell that plants have taught this medicine to them. In the book "The cosmic Serpent" [J10], Jeremy Narby proposes that shamans communicate directly with the DNA of the plant and that bio-photons might be essential for these communications.

2. Are "skin senses" in preferred role in inter-species communications

The general model of sensory organs [K4] provides further pieces to the model. The basic hypothesis is that sensory organs are the seats of primary qualia and that the experiencing self corresponds to the magnetic body of size much larger than the material body. The feedback via brain to the sensory organs is assumed to make possible the active processing of the sensory input to sensory percept.

The model introduced also a division of sense to "brain" and "skin" senses motivated by the following observation. During the development of vertebrate embryo three basic types of cells are formed: ectoderm giving rise to skin and central nervous system, endoterm giving rise to many internal organs, and mesoderm giving rise to muscles, skeleton, connective tissue, ... What is remarkable is that the ectoderm giving rise to brain, spinal cord, eye, .. suffers an inversion during the formation of the neural tube. The wall of the neural tube is inverted so that inside becomes outside and vice versa. This explains the strange finding that the eyes of vertebrates are inversion of the eyes of invertebrates and apparently look very awkward from engineering view point.

This leads to the distinction between "brain senses" and "skin senses".

1. For "brain senses" (vision, olfaction) the back projection by the telepathic sharing of mental images from brain to sensory organs allows to build sensory percept as a caricature.

2. For "skin senses" (with hearing included) the entanglement of the sensory organ with brain is replaced with the entanglement with external world, and can thus give rise to remote sensing

based on the sharing of mental images using time mirror mechanism. This remote viewing need not be conscious-to-us except in special situations, say during hypnosis.

If this distinction is sensible, skin senses would be in a special role as far interspecies communications are considered.

5.5.3 A model for insect-plant communications

Callahan has made very important discoveries related to the olfaction of insects and insect-plant interaction. Callahan's work [I20, I8] demonstrates that the insect olfaction is based on infrared light generated by the odorant molecules interacting with the antennae of the insects. This might be true for our olfaction too. Callahan has also shown plants communicate with insects by generating infrared light [I8]. This finding conforms with the findings of Albrecht-Buehler [I15] showing that all cells contain microtubular structures acting as receiving antennae for infrared light. Furthermore, plants suffering from de-nutrition are found more easily by insects than healthy plants.

These findings encourage to consider the following mechanism for insect-plant communications. This mechanism could apply also to the plant-human interactions.

1. Insects generate infrared MEs propagating like massless particles inside low frequency negative energy MEs acting as bridges quantum entangling the plant and the insect. Both classical communications by positive energy IR MEs and quantum communications by negative energy IR MEs are in principle possible.

2. In the case that plant suffers from de-nutrition, it can gain metabolic energy by sending negative energy MEs received by insect. This gives for plant metabolic energy and at the same time generates the quantum entanglement bridge making it possible for the insect to find the plant. The same mechanism explains also the episodal memory feats of synesthetes: due to the over-activity of subcortical parts of brain the neurons of the left cortex suffer starvation and generate negative energy MEs providing them metabolic energy and simultaneously entangling them with the geometric past so that episodal memories result. Also the life review of NDE experiences could be a by-product of the neuronal starvation.

5.5.4 Are human-plant communications possible?

Concerning the ideas about human plant-communications, I am grateful for Peter Hageman [I1] for inspiring discussions. Hageman claims to be able to express "plant language" by "dancing" in meditative state, and has made detailed maps of motor expressions of plant mental images and proposed a detailed taxonomy about plant language, and also developed refined ideas such as the notion of dialect. It is easy to debunk Hageman's views since the interpretations necessarily involve a lot of subjective elements. One could imagine a scientific testing of Hageman's claims by checking whether the vocabulary defined by the bodily expressions is invariant of a given plant. Similar testing is used in the case of motor synesthesia.

The common portions of the memone and genome of plant and human would in principle make possible this communication. Thus the memes and genes, which plant can communicate to humans are those which are expressed also by some cells of the human body. Since plants do not have a central nervous system, it seems that the communicable memes and genes should correspond to rather primitive expressions common to us and animals. Obviously, the memes and genes responsible for the body language is a good first guess in this respect. Note however that genes could define lower level language and could also be communicated in this manner.

This body language representation brings in mind motor synesthesia in which sensory input from some sensory modality is expressed as motor activity. The memes and genes expressible by body language could of course be expressed also by using ordinary language. The universality of the memetic and genetic codes is indeed consistent even with the transformation of communications to linguistic expressions.

5.6 Intronic Portions Of Genome Code For RNA: For What Purpose?

The last issue of [I9] contains an article about the discovery that only roughly one half of DNA expresses itself as amino-acid sequences. A detailed summary of the results has been published

in Nature [I4]. The Encyclopedia of DNA Elements (ENCODE) project has quantified RNA transcription patterns and found that while the "standard" RNA copy of a gene gets translated into a protein as expected, for each copy of a gene cells also make RNA copies of many other sections of DNA. In particular, intron portions ("junk DNA", the portion of which increases as one climbs up in evolutionary hierarchy) are transcribed to RNA in large amounts. What is also interesting that the RNA fragments correspond to pieces from several genes which raises the question whether there is some fundamental unit smaller than gene.

None of the extra RNA fragments gets translated into proteins, so the race is on to discover just what their function is. TGD proposal is that the RNA gets braided and performs a lot of topological quantum computation [K16]. Topologically quantum computing RNA fits nicely with replicating number theoretic braids associated with light-like orbits of partonic 2-surfaces and with their spatial "printed text" representations as linked and knotted partonic 2-surfaces giving braids as a special case. An interesting question is how printing and reading could take place. Is it something comparable to what occurs when we read consciously? Is the biological portion of our conscious life identifiable with this reading process accompanied by copying by cell replication and as secondary printing using amino-acid sequences?

This picture conforms with TGD view about pre-biotic evolution. Plasmoids [I18], which are known to share many basic characteristics assigned with life, came first: high temperatures are not a problem in TGD Universe since given frequency corresponds to energy above thermal energy for large enough value of $\hbar$ [K18]. Plasmoids were followed by RNA, and DNA and amino-acid sequences emerged only after the fusion of 1- and 2-letter codes fusing to the recent 3-letter code. The cross like structure of tRNA molecules carries clear signatures supporting this vision. RNA would be still responsible for roughly half of intracellular life and perhaps for the core of "intelligent life".

I have also proposed that this expression uses memetic code which would correspond to Mersenne $M_{127} = 2^{127} - 1$ with 2^{126} codons whereas ordinary genetic code would correspond to $M_7 = 2^7 - 1$ with 2^6 codons. Memetic codons in DNA representations would consist of sequences of 21 ordinary codons. Also representations in terms of field patterns with duration of.1 seconds (secondary p-adic time scale associated with M_{127} defining a fundamental bio-rhythm) can be considered.

A hypothesis worth of killing would be that the DNA coding for RNA has memetic codons scattered around genome as basic units. It is interesting to see whether the structure of DNA could give any hints that memetic codon appears as a basic unit.

1. In a "relaxed" double-helical segment of DNA, the two strands twist [I6] around the helical axis once every 10.4 base pairs of sequence. 21 genetic codons correspond 63 base pairs whereas 6 full twists would correspond to 62.4 base pairs.

2. Nucleosomes [I5] are fundamental repeating units in eukaryotic chromatin [I2] possessing what is known as 10 nm beads-on-string structure. They repeat roughly every 200 base pairs: integer number of genetic codons would suggest 201 base pairs. 3 memetic codons makes 189 base pairs. Could this mean that only a fraction $p \sim 12/201$, which happens to be of same order of magnitude as the portion of introns in human genome, consists of ordinary codons? Inside nucleosomes the distance between neighboring contacts between histone and DNA is about 10 nm, the scale $L_e(151)$ associated with the Gaussian Mersenne $(1+i)^{151} - 1$ characterizing also cell membrane thickness and the size of nucleosomes. This length corresponds to 10 codons so that there would be two contacts per single memetic codon in a reasonable approximation. In the example of Wikipedia [I5] nucleosome corresponds to about 146=126+20 base pairs: 147 base pairs would make 2 memetic codons and 7 genetic codons. The remaining 54 base pairs between histone units + 3 ordinary codons from histone unit would make single memetic codon. That only single memetic codon is between histone units and part of the memetic codon overlaps with histone containing unit conforms with the finding that chromatin accessibility and histone modification patterns are highly predictive of both the presence and activity of transcription start sites. This would leave 4 genetic codons and 201 base pairs could decompose as memetic codon+2 genetic codons+memetic codon+2 genetic codons. The simplest possibility is however that memetic codons are between histone units and histone units consist of genetic codons. Note that memetic codons could be transcribed without the straightening of histone unit occurring during the transcription leading to protein coding.

6 Corals And Men

The comparison of the human genome with that of invertebrates during great genome project yielded some surprises [I10, I16]. In particular, it was found that human genome possesses 223 genes which are lacking from the genome of the model invertebrates. Even more, a considerable fraction of these genes seems to be lacking even from the genome of other vertebrates and should have thus appeared relatively recently to the human genome. This led to the TGD inspired proposal that human genome has been subject to genetic engineering relatively lately with an even more crazy sounding proposal that intelligent intra-terrestrial life forms are responsible for this engineering (see [K1, K2]). Needless to say, this hypothesis is absolute non-sense unless one takes completely seriously the notion of many-sheeted space-time and related ideas (see **Fig.** `http://tgdtheory.fi/appfigures/manysheeted.jpg` or **Fig.** 9 in the appendix of this book).

Towards the end of year 2003 a new sensational result was published [I14]. The genome of certain coral, Cnidarian Acropora millepora, was compared with the genomes of fly (Drosophila melanogaster), nematode worm (Caenorhabditis elegans) and human genome. For simplicity, I shall refer to these animals as corals, flies, and worms in the sequel.

Before going to results, it deserves to mention that the phylum Cnidaria [I3] is a major group of invertebrates that includes the sea anemones, corals, jellyfishes, hydroids. Radial symmetry is characteristic for these animals whereas more advance animals have bilateral symmetry. All cnidarians are carnivors. The name cnidaria comes from "cnida", stinging capsules used to deliver toxin, to stick to a prey, or to entangle with an object. Cnidarians are believed to branch from the evolutionary tree much before flies, worms and vertebrates, and the expectation was that its genome should give information about the genome of the common predecessor of metazoans ("animals"). Obviously, coral genome should possess genome having much stronger resemblance to fly and worm genomes than that of vertebrates.

The astonishing result was that the genome of coral resembles human genome to a surprisingly high degree but much less the genomes of the fly D. melanogaster and nematode worm C. elegans. The conclusion of the authors is that the genes usually regarded as vertebrate inventions have been possessed by the believed-to-exist common predecessor of all animals, and that invertebrates have lost most of the genes common to human and coral during their evolution.

These strange conclusions are based on the existing wisdom about the evolution of animals and about the role of genome, and thus also challenge this wisdom.

1. The mutation rate of flies and nematode worms is high but it is not known whether genes can disappear. One can of course criticize the conclusion. Is the loss of genes, kind of "de-evolution", really a good survival strategy? Could it be that the gene loss hypothesis is dictated by some hidden assumptions? One can imagine several assumptions of this kind.

   (a) It makes sense to speak about continuously growing evolutionary tree from which corals branch before the emergence of the flat worms believed to be common predecessors of vertebrates, flies, worms, and other invertebrate phyla.

   (b) There are no of interventions from outside to the genetic evolution (genetic engineering by some more advanced life forms).

   (c) There is no horizontal gene transfer between corals and vertebrates (say fishes) living belonging to the ecosystems containing corals.

2. Corals possess relatively few tissue types. Coral is in fact the the most primitive animal possessing neurons. The neurons are organized into a homogenous neural net like structure. Therefore the presence of genes of vertebrates with highly developed nervous systems in the coral genome looks paradoxical. Could genes have some unknown functions besides those related to the coding of "hardware" ?

6.1 Why Corals And Vertebrates Should Have Common Genes?

The fact that corals and higher vertebrates have a lot of common genes, must have some sensible explanation. If flies and worms have never possessed genes common to vertebrates and corals, there must be something common to corals and vertebrates but not to higher vertebrates and flies, worms, perhaps all other invertebrates than corals.

1. Corals (polyps) are simple multi-cellulars consisting of two cell layers, epidermis and gastrodermis. The non-tissue layer between them is called the mesoglea. Corals are in a complex interaction with an ecosystem involving also vertebrates. Interestingly, corals are the only animals living in symbiosis with plants (dino-flagellates, which are simple mono-cellulars) providing them metabolic energy produced in photosynthesis. Social life requires communications and language of some kind, and this might explain why corals possess so complex genome.

2. Corals form large populations, coral reefs. One could regard coral reef as a super-organism, with corals taking the role of the cell in ordinary organisms. Some corals could even define super-neurons of the super-neural system of this super-organism. This could resolve the paradox caused by the simplicity of the nervous system of the coral itself viz. the complexity of the coral genome: complexity would reside in the connections and communications of the super-neural system.

3. Rather remarkably, both vertebrates and corals possess skeletons having calcium as a basic building brick. Coral skeleton consists of calcium carbonate $CaCO_3$ or limestone (calcium carbonate plus sediment). Usually the basic function of skeleton is thought to be that of a mere supporter but it might have also other functions as following arguments suggest.

The function in question could be that of antenna.

1. Ca_{++} waves are known to play key role in the functioning of the nervous system [J11]. The finding, which led to the notion of magnetic body crucial for the functioning of living systems in TGD Universe, was that irradiation of living matter at the harmonics of Ca_{++} cyclotron frequency in Earth's magnetic field has effects in living matter [J6] and that these effects can be best understood if living cells are macroscopic quantum systems.

2. What makes Ca_{++} (and Mg_{++}) ions so special that they are bosons and can therefore form super-conducting Bose-Einstein condensates at the magnetic flux tubes of (say) Earth's magnetic field. The dropping of Ca_{++} ions from smaller space-time sheets to the magnetic flux tubes generates cyclotron radiation at the harmonics of the cyclotron frequency, and this mechanism is crucial for the generation of EEG.

   In many-sheeted space-time particles topologically condense at all space-time sheets having projection to given region of space-time so that this option makes sense only near the boundaries of space-time sheet of a given system. Also p-adic phase transition increasing the size of the space-time sheet could take place and the liberated energy would correspond to the reduction of zero point kinetic energy. Particles could be transferred from a portion of magnetic flux tube portion to another one with different value of magnetic field and possibly also of Planck constant h_{eff} so that cyclotron energy would be liberated.

3. The fact that EEG is associated with vertebrate neural systems and corals representing the simplest life forms having neurons, suggests that communications could be on Ca_{++} cyclotron radiation and be between neural systems of corals and vertebrates living around it. The calcium containing skeletons possessed by both vertebrates and corals could serve as antennae allowing to generate field patterns representing the communicated signals strong enough to make possible communications with say fishes.

These observations can be combined with the earlier vision about the role of genome.

1. TGD based view for genetic code predicts that genes are not expressed only chemically but also in terms of field patterns representing one particular expression of the universal language defined by genes and memes, the latter being represented by the intronic portions of the genome. The hypothesis is that common memes and genes make possible even inter-species communications. Thus DNA would not characterized so much species than communications in the ecosystem in which the species.

2. The proposed electromagnetic realization of the language defined by DNA and utilizing genetic code is as transitions at the harmonics of cyclotron frequency of Ca_{++} ions. The time

$\tau_c = 1/f_c$ defined by cyclotron frequency $f_c = 15$ Hz of Ca_{++} ion in the magnetic field of $B_e = .5$ Gauss, defines the duration of the code word consisting of 6 bits. Bit "1" can be realized as a pulse of duration about $\tau_c/6$. Second realization if the codeword is as superpositions 6 lowest harmonics of cyclotron frequency generated when Ca_{++} ions drop to larger space-time sheets (bit "1" corresponds to harmonic with Fourier amplitude above critical intensity).

The combination of these ideas leads to the following vision about why corals and vertebrates possess common genes.

1. Linguistic communications based on electromagnetic field patterns are crucial for the functioning of the complex ecosystem formed by corals and various species belonging to it. In particular, fishes are vertebrates so that it would be advantageous for the coral to possess a considerable fraction of vertebrate genome responsible for the universal language.

2. The basic difference between humans and corals is due to the intronic portion of the genome, which dominates human genome and defines higher level linguistic structure based on menetic codewords with duration of.1 seconds and having 126 (or 127) bits. One could test whether fishes and other vertebrates living in the coral environment share identical genes with the coral.

6.2 Did Corals And Vertebrates Receive Their Common Genes Via Horizontal Transfer?

One can imagine several alternative models for how corals and vertebrates received their common genes. These "language genes" could have been "invented" by either corals or vertebrates and transferred horizontally from corals to vertebrates or vice versa. The most general view is that the system inventing the "language genes" was the entire system coral + vertebrates (possessing magnetic body), and the transfer of genes occurred in both directions.

Many-sheeted space-time allows to consider the possibility that the genes were transferred along magnetic flux tubes connecting the many-sheeted DNAs of corals and vertebrates involved (recall that many-sheeted DNA involves hierarchy of magnetic flux tubes with magnetic field strengths scaled by powers of two). The large sized magnetic body of the coral might have played essential role in this transfer. If one sees the magnetic body associated with corals and associated life-forms as a single conscious organism, this hypothesis looks natural.

6.2.1 Horizontal transfer of genes from vertebrates to corals?

Perhaps the simplest hypothesis is that the genes of vertebrates not possessed by invertebrates like flies and worms, appeared in the genome of corals, only after they were needed, that is during the co-existence with vertebrates as parts of same ecosystem.

The simplest possibility is a horizontal gene transfer from vertebrates to corals, so that vertebrate genes could be seen as genuine vertebrate inventions as usually believed. Since corals possess calcium backbone, it is indeed sensible to transfer "language genes" to corals. The transfer of "language genes" could have occurred also to other organisms but, by the absence of Ca containing skeleton, the transferred "language genes" would have been useless for them. The genes coding for the calcium containing skeleton of coral must have been present from the beginning, and for this option must be independent of "language genes".

6.2.2 Horizontal transfer of genes from corals to vertebrates?

The horizontal transfer of genes could have also occurred from corals to vertebrates so that vertebrate genes could be seen as coral inventions. This makes sense if the common genes were primarily involved with the communications between corals and vertebrates so that the co-evolution of vertebrates and corals was basically evolution of communications using genetic code. Also the vertebrate genes responsible for coding Ca_{++} based structures like bones should have counterparts in the genome of the coral. If corals possessed neurons from the beginning, the genes allowing to differentiate the cell to a neuron could have been transferred, not only to vertebrates but also to

the invertebrates with nervous system, and corals could have functioned as a kind of organizing centers of evolution.

6.3 What Happened In Cambrain Explosion?

The most science fictive hypothesis is that the vertebrate genes were possessed by corals from the beginning. Ironically, this is the hypothesis forced also by the standard view about evolutionary tree. Since these genes had no function at the time when even flat forms preceding flies, worms, and vertebrates were still absent, the only reasonable conclusion seems to be that some advanced life forms intervened the terrestrial evolution somehow.

There are two basic options.

"Language" genes could have appeared to the genome of coral by some kind of genetic engineering. Since corals are the simplest organisms possessing neurons, also the genes responsible for the differentiation of cells to neurons could have emerged during this intervention.

2. Corals themselves could have been these extra- or intra-terrestrial life forms possessing these genes.

Corals would have served as kind of gene banks and "language" genes would have been transferred horizontally to the genomes of vertebrates during the evolution of the ecological co-existence with vertebrates such as fishes. "Neuronal genes" would have been transferred also to other phyla than future vertebrates.

Obviously, this intervention could have have induced profound changes, entirely new phyla could have emerged as a result of these genetic modifications by horizontal gene transfer. Although this scenario sounds un-necessarily science-fictive, it deserves a serious consideration since, as will be found, it could allow to resolve the puzzles related to the Cambrian explosion.

This option can be tested by looking whether corals possess genes possessed by higher vertebrates like humans but not by any of the vertebrates, which have possibly been members of same ecosystem with corals. One could also try to test the hypothesis that coral has possessed "language genes" already at Cambrian period. This could be the case, if "language genes" and genes coding for Ca^{++} body of coral belong to the same gene cluster.

6.3.1 Did Cambrian explosion involve the intervention of intra- or extraterrestrial life forms?

According to the fossil records, multicellular fauna emerged suddenly in the so called Cambrian explosion [I21]. Already Darwin realized that the absence of fossils from pre-Cambrian era posed serious challenges for the idea about gradual evolution in which evolutionary tree develops gradually new branches. The problems are following.

1. Why did the Cambrian explosion occur so late? This problem is discussed in the chapter [K19].

2. Why no fossils of the pre-Cambrian predecessors of the Cambrian fauna have been found? The obvious explanation is that the faunas of pre-Cambrian era did not contain hard parts and thus yielded no fossils. During last decades two pre-Cambrian faunas have been however found. Edicaria are multi-cellulars with a pan-cake like shape whereas Tommotian fauna consists of simple cup and cap like multi-cellulars [I21]. Contrary to what one might expect, there is however no continuous evolution of these faunas to the Cambrian fauna and it seems that Cambrian explosion suddenly generated the predecessors of the recent day fauna.

3. The third problem which Darwin could not yet face emerged much later. The Burgess Shale fauna found 1909 by Charles Doolittle Wallcott challenged even the very notion of the evolutionary tree (for an excellent popular discussion of implications of Burgess Shale fauna see the book "Wonderful Life" of Stephen Jay Gould [I21]). It was found that several phyla of animal kingdom (kingdoms consist of phyla), which do not exist today, emerged in Cambrian explosion and then disappeared. The usual view about evolution is that it creates complexity from simplicity rather than vice versa. Obviously, just the opposite occurred in

Cambrian explosion. Even more remarkably, not a single phylum has appeared to the fauna after the Cambrian explosion.

All these problems would have a nice resolution if Cambrian explosion were due to the intervention of intra- or extra-terrestrial life forms. Suppose that life evolved in intra-terrestrial conditions, in the safe womb of Mother Gaia, one might say. In this frame of mind Cambrian explosion could be seen as a moment when Mother Gaia gave birth to new phyla preceding the recent day fauna. After the moment of birth these children of Mother Gaia had to survive by their own in the harsh Cambrian environment and many of them did not. This would elegantly explain why Cambrian explosion was followed by a strong extinction of phyla.

The metaphor about mother Gaia giving birth to life forms could have at least two meanings.

1. The most concrete meaning is as the emergence of these life-forms from intra-terrestrial conditions to the surface of Earth: many-sheeted space-time might make this Jules Verne like travel possible. Corals might have been these intra-terrestrial life-forms realizing genetic code in terms of field patterns using calcium containing skeletons as antennae, and being the first animals possessing primitive nervous system and ability to realize genetic code electromagnetically in EEG frequency range. After this the horizontal transfer of intra-terrestrial genes would have stimulated the evolution of vertebrates from more primitive life forms and have generated various phyla.

   Although corals are simple life forms, coral populations are not, and they could be seen as super-organisms with cells being replaced by double cell layers, corals. If this interpretation is accepted, coral populations represent "alien" life forms at a higher level in the evolutionary hierarchy. This interpretation would make natural the identification of corals as organizing centers of evolution. The radial geometry of corals would nicely symbolize their role as evolutionary organization centers.

2. A less concrete realization of the metaphor would be as intra-terrestrial genetic engineering involving the insertion of packets of introns and genes (software plus necessary hardware) to the DNAs of already existing life forms. Of course, option

3. could be seen as are representing a special manner to perform this genetic engineering by utilizing a life form possessing the genes possessed also by the highest vertebrates and radiating them horizontally.

6.3.2 Variants about the genetic engineering theme

The idea about genetic engineering performed by intra- and/or extraterrestrials could have been realized in several manners.

1. The first vertebrates got their vertebrate genes from intra- or extraterrestrials, and these genes were horizontally transferred to corals living in symbiosis with them. This genetic intervention should have occurred after the Cambrian explosion if it turns out that the nearest predecessors of vertebrates do not possess the common genes.

2. The transfer of genes was from corals to vertebrates and corals received their vertebrate genes in Cambrian explosion from some highly advanced life form, or corals themselves were this advanced life form. During the gradually evolving co-existence with simpler life forms these genes were gradually horizontally transferred to the predecessors of recent day vertebrates.

3. Corals and the predecessors of the vertebrates living in symbiosis with them got their common vertebrate genes simultaneously from outside. Note however that they could have been inventions made by the entire ecosystem in question.

6.4 What Ontogeny Recapitulates Phylogeny Principle Means At The Level Of DNA?

The idea about corals as a gene bank could resolve some other mysteries related to the genome.

1. The revised view about the evolution of organisms possessing nervous systems would not rely so much on mutations but already existing and possibly conserved gene and meme banks inherited from corals. Only a fraction of these genes would be in use and evolutionary pressures would determine which portion of these memes/genes is activated. Mutations might only have a secondary and mostly harmful role in the evolution.

2. Introns are the basic candidate for the gene reservoir. This conforms with the idea that introns code for memes since it is plausible that the activation of memetic programs becomes possible only when genes coding for the needed hardware are activated.

The quite recent finding [J15] that the removal of massive portions of the conserved part of the genome has no apparent effect on the organism's functioning supports the idea about genetic repertoire. It is usually believed that conserved regions of genome have some function. Many of these conserved regions are in the "junk" portion of DNA and do not code for proteins. They could however still have some function and in some cases the conserved regions indeed seem to affect the expression of the nearby genes. To identify the function of the highly conserved intronic regions the team of Edward Rubin at the Lawrence Berkeley National Laboratory deleted two huge regions of intronic DNA from mice containing nearly 1000 highly conserved sequences shared by human and mice. One of the chunks was 1.6 million DNA bases long, the other one was over.8 million bases long. The unexpected result was that the genetically modified mice were virtually indistinguishable from normal mice in every characteristic they measured, including growth, metabolic functions, lifespan and overall development.

The result could be understood if the intronic portion of both mice and men derives from a very early period of evolution, perhaps the evolution leading to corals (where-ever it occurred). These conserved genes would have developed to their stable forms during the evolution leading to corals and represent a repertoire of functions waiting for their activation. One could understand the instantaneous popping up of highly developed biological functions, which is often used as a counter argument against the view that the evolution is made possible by random mutations. The interpretation also conforms with the idea that at least part of introns code for memes. In the case of mouse memone would not be yet of much use since speech organs and culture are lacking. The comparison of the introns of man and apes to see whether memes having some function related to language are active in humans but not in apes could serve as a test of this prediction.

Ontogeny recapitulates phylogeny principle would be realized as a gradual shift of the activated portion of the memone and genome during the development of the embryo. Also the memes and genes still waiting circumstances allowing for their expression would be present in the DNA. It might be possible to create artificially earlier evolutionary forms of a given organism and evolution might be studied in the laboratory. The partial transmutation of the morphologies of two different organisms to each other might be possible if the organisms possess common portions of the conserved genome.

If also the future phylogeny is coded to the recent DNA, both mice and men would possess enormous evolutionary potential (frog to prince effect!). Perhaps the explosive cultural evolution of civilization during last centuries has been accompanied by a corresponding shift in the activated portion of memone. This shift could also occur during the lifespan of individual and the idea about personal growth would have a genetic justification. The evolutionary potential might be some day be utilized by the artificial activation of memone and genome. Of course, the activation of higher memetic programs would be possible only if the genes coding for the needed hardware are also present and activated. Again the computer metaphor would work: I have used only a minor portion of the potential of my text processing program to write this piece of text.

6.4.1 Gene activation by electrostatic fields?

The basic question concerns the method of activation. The discovery of chemists Guido Ebner and Guido Schuerch [B1] , [J2] raises the hope that these ideas might be more than over-active imagination and their work also provides a concrete proposal for the activation mechanism. These findings are briefly described in the article of Hardmuth Mueller [B1] who proposes quite different explanation for the strange findings. Ebner and Schuerch studied the effect of electrostatic fields on the growth and morphogenesis of various organisms. Germ, seeds, or eggs were placed between

conducting plates creating an electric field in the range.5-2 kV/m: note that the Earth's electric field is in the range $.1 - 4$ kV/m and of the same order of magnitude.

The outcome was rather surprising and in the year 1989 their employer Ciba Geigy (now Novaris) applied for a patent "Method of enhanced fish breeding" [J2] for what is called Ciba Geigy effect. The researchers describe how fishes (trouts) develop and grow much better, if their eggs have been conditioned in an electrostatic field. The researchers report [J2] that also the morphology of the fishes was altered to what seems to represent an ancient evolutionary form: this was not mentioned in the patent.

The chemists founded their own Institute of Pharmaceutical Research near Basel, where Guido Ebner applied for another very detailed patent, which was never granted (it is not difficult to guess the reasons why!). In the patent he describes the effect of electrostatic fields on several life forms (cress, wheat, corn, fern, micro-organisms, bacteria) in their early stage of development. A clear change in the morphogenesis was observed. For instance, in one example fern had all sort of leaves in single plant apparently providing a series of snapshots about the evolution of the plant. The evolutionary age of the first leaf appeared to be about 300 million years whereas the last grown-up leaf looked close to its recent form.

If one takes these finding seriously, one must consider the possibility that the exposure to an electrostatic field can activate passive genes and change the gene expression so that older morphologies are expressed. The activation of not yet existing morphologies is probably more difficult since strong consistency conditions must be satisfied (activation of program requires activation of a proper hardware).

It is known that the developing embryo has an electric field along the head-tail axis and that this field plays an important role in the control of growth. These fields are much weaker than the fields used in the experiment. p-Adic length scale hierarchy however predicts an entire hierarchy of electric fields and living matter is indeed known to be full of electret structures. The strength of the electric field in some p-adic length scale related to DNA might somehow serve as the selector of the evolutionary age. The recapitulation of phylogeny during the ontogeny could mean a gradual shift of the activated part of the memone and be controlled by the gradually evolving electric field strength.

The finding that led Ebner to his discovery was that it was possible to "wake up" ancient bacteria by the exposure to an electrostatic field. This would suggest that in the case of primitive life forms like bacteria the electric field strength of Earth controls the state of bacterium whereas in higher life forms endogenous electric fields have taken the role of Earth's electric field.

6.4.2 Electric fields and healing

Wound healing is very much like morphogenesis and already Becker discovered that electric field induces a healing of wounds [J4]. More recent studies of the effects of electric fields on healing and on embryos are discussed [J7] and one can find useful quantitave information from this article. The typical strengths of the electric fields appearing in organisms are in the range.01-.1 kV/m: the upper bound of this field is the lower limit of Earth's recent electric field strength and considerably below the field yielding Ciba Ceigy effect. There are however also much stronger fields present: the voltage over epithelium (double cell layer) is in the range 30-50 mV and would make 6-10 kV if the thickness of epithelium were 5 μm. In the 1950s it was discovered that the direction of external electric field determines whether a flatworm which has been cut in two pieces develops head or tail. A natural voltage gradient exists between the severed worm's tail and the place where its head once was. A naturally occurring electric field of strength 40 V/m has been found to play a vital role in the wound healing in the cornea of rats: it is found that the cells divide in the plane orthogonal to the field and pushing new cells to the wound.

A further finding reported in [J7] was that a voltage of 2 mV over cell diameter (1 kV/m if cell radius is 2 μm and in the same range as the field applied in Ciba Geigy effect) alters the front back orientation of the neuroblast. The proposed interpretation of these findings is that electric field provides only directional information whereas the findings of Ebner suggest that much more profound meme/gene level effects might be involved. For instance, one can ask whether the exposure to an appropriate external electric field could induce the return of a differentiated cell to the stem cell stage realized as a shift in the activated portions of memone and genome.

6.4.3 Generalized four-wave mechanism and the concrete mechanism of gene activation by static electric fields

Concerning the concrete mechanism behind the activation there are several constraints. The activation mechanism must be localized at the level of DNA. On the other hand, a given region of DNA must activate only in an electric field of a particular strength coding for its evolutionary age. The basic finding of Ebner about "wake-up" of ancient bacteria in a static electric field suggests that the activation must be a kind of "wake-up" process for an appropriate part of DNA. "Wake-up" corresponds to the generation of self-organization pattern getting its metabolic energy by sending negative energy photons absorbed by some system with corresponding excitation energy. This mechanism is indeed non-local since only the strength of the electric field matters. The transfer of an electron in an electric field through a distance not larger than the size of nucleus by absorbing negative energy photon is a good candidate here.

The generalized four-wave mechanism for remote metabolism requires the existence of generalized standing waves taking care of themselves by sending negative energy (phase conjugate) photons at the energy defined by the frequency of wave. The ideal situation corresponds to a dispersion relation for which the frequency of the oscillation does not depend on wave vector at all: plasma oscillations satisfy this conditions. In the chapter [K7] a model for the coherent electric dipole oscillations was constructed. In this case the frequency depends only on the angle between the wave vector and the direction of the electric field. Same applies to magnetostatic oscillations [D1] for which the Larmor frequency of electron gives the maximum value of frequency. The frequency is constant for effectively 2-dimensional oscillation patterns with wave vectors at the surface of a cone and periodically recurring oscillation patterns able to represent simple 2-dimensional self-sustaining mental images become possible.

Quantitative estimate support this model. In a field of.5-2 keV/m the electron gains a kinetic energy of $5-20$ μeV while travelling a distance of 10 nm corresponding to the thickness of cellular membrane. This corresponds to photons with microwave frequencies in the range .12-.48 GHz and wave lengths in the region.6-2.5 m so that the energy can be sucked from quite large spatial volume by inducing transfer of electron through this distance without a gain of kinetic energy. These frequencies corresponds to the nanosecond scale assigned with the coherent electric dipole oscillations whose importance was first realized by Fröhlich [I17]. In the above mentioned model of coherent dipole oscillations as analogs of magnetostatic oscillations these frequencies correspond the p-adic length scale $L(151) = 10$ nm associated with cell membrane thickness which is of central importance in the coiling hierarchy of DNA.

Also magnetostatic waves in a magnetic field associated with some level of the hierarchy of DNA space-time sheets could be in question. In this case the Larmor frequency of electron defines the maximal oscillation frequency. From the assignment of $L(169) \simeq 5$ μm to the Earth's magnetic field, the length scale associated with the needed field varying in the range $.05-.1$ Tesla is in the range.1-.2 μm.

6.5 Where Did Those 223 Genes Pop Up?

The reports of the Public Consortium about human genome in Nature, Feb 15, 2001 [I10] and of Celera Genomics in Science of Feb 16th, 2001, [I16] contained two big surprises.

6.5.1 Are we really so near to fruit flies?

The first astonishing discovery was that the amount of human genome differs relatively little from those of lower organisms: we have only about 30, 000 genes, little more than twice the number 13, 601 of genes for fruit fly. This paradoxical finding forces to think that our genome is not solely responsible for what we are and that the intronic portion of DNA (only about 1 per cent codes of human DNA codes or amino-acid sequences), is not "junk DNA", but contains important biological information and expresses it non-chemically.

In TGD Universe introns would express memes as the classical field patterns associated with MEs ("topological light rays") responsible for the basic expressions of language understood in an extremely general sense. This language includes body language and even cellular signalling, and could quite well make possible (not necessarily conscious) interspecies communications based on memes expressed by communicating species and forming a common vocabulary. All eukaryotes

(cells with nuclei), even bacteria, would possess part of the vocabulary of this universal language. The memetic code word is predicted to consist of a sequence of 21 DNA triplets and carries 126 bits of information instead of 6 bits of genetic code. Of course, also genes are expressed in terms of MEs and define a lower level language.

In this framework the actual role of DNA can be understood using the computer analogy. Memes represent the program modules written using the programming language defined by the memetic code, and realized in terms of the field patterns associated with MEs. Genes represent the necessary hardware needed to realize these programs. System builds only the hardware needed, that is cell expresses only part of the genome. DNA engineering requires besides the addition of the new programs (memes, introns) also the insertion of the necessary hardware (new genes). Memes and corresponding genes should have very intimate relationship. In this conceptual framework the standard view is wrong since it identifies the build-up of a new hardware as the sole activity at the DNA level. This would be like identifying the addition of a net card to a computer as the fundamental activity related with computers.

6.5.2 The head-scratching discovery

The "head-scratching discovery" by the public consortium, as Science termed it, came when the genome was compared with the genomes of our predecessors. It was found that human genome contains 223 genes not possessed by invertebrates. Contrary to what one might expect, these 223 genes could make an enormous difference. The reason is that this number is more than two thirds of the number of the 300 genes differentiating between humans and chimpanzees so that these genes could be the main determinant of the dramatic difference between humans and chimpanzees in standard genetics.

Of course, in TGD framework the most important differences would probably relate to the intronic portion of the DNA responsible for language. Dramatic differences between our intronic DNA that of our invertebrate and perhaps even vertebrate predecessors, in sharp conflict with the idea of continuous evolution, should be discovered.

6.5.3 Are the enigmatic genes a horizontal gene transfer from bacteria?

Biologists can explain the presence of the enigmatic genes only by a "rather recent horizontal transfer from bacteria". Here "rather recent" refers to the evolutionary time scale.

This explanation can be challenged on various grounds.

1. The simplest working hypothesis is that the transfer from bacteria is a probabilistic process. The problem is however why the horizontal transfer did not occur to the genomes of other vertebrates and invertebrates and gradually through the whole evolution. One could argue that something characteristic to the vertebrate genome should have made this process possible. In TGD framework one could imagine that the intronic portion of the vertebrate genome could have contained something which made the transfer possible: a common part of memone with the bacteria involved and making possible language based communications ("language" understood in a generalized sense) at DNA level perhaps?

2. The enigmatic genes are involved with important physiological functions. In particular, they are responsible for important neurological enzymes which stem from mitochondria having its own genome. According to my non-professional interpretation this statement means that also mitochondrial genome contains these enigmatic genes. Thus both mitochondrial and nuclear genomes would have been altered by this horizontal transfer from bacteria. Simultaneous double horizontal transfer does not however look a probable event.

3. Only 113 of the 223 enigmatic genes are widespread in bacteria: it would be easier to believe in the horizontal transfer if all of them were widespread. These 113 widely occurring genes are not encountered in invertebrates at all. As a matter fact, this finding suggests that the transfer occurred from the vertebrate genome to the bacterial one and only partially, rather than vice versa. The analysis of proteins expressed by the enigmatic genes demonstrated that out of 35 identified, only 10 had counterparts in other vertebrates. 25 of them were unique to humans. This suggests that a considerable part of the horizontal transfer has

occurred relatively recently and together with associated introns might even distinguish us from chimpanzees.

6.5.4 Horizontal transfer as DNA engineering?

The objections against the horizontal transfer from bacteria force to consider seriously the possibility that the horizontal transfer represents an intentional DNA engineering, both memetic and genetic. The most important transfer should have been to the intronic part of the DNA. The addition of memes would be like adding a new program to a computer. The addition of genes would be like adding a new hardware (say net card or data cable) required by the program to run. The comparison of the intronic portions of DNA of humans and lower vertebrates might thus lead to futher "head-scratching" discoveries. The data are consistent with the assumption that genetic/memetic engineering activities have occurred in several steps during the evolution of the vertebrates although a considerable portion of the enigmatic genes and associated introns, perhaps even two thirds, have been "injected as a single dose".

The evolution of the hominides in Africa had a stagnation period of about 1.5 million years as demonstrated by the study of the ancient stone tools. Then, for about 50 thousand years ago, a sudden jump to creativity occurred. The first ornaments appeared meaning that hominides had become artists and started to express their position in the social hierarchy by clothing and ornaments. This signals about development of highly refined social structures. A general belief is that also language began to develop rapidly and made possible a cumulation of knowledge. It seems that modern human was born and started to migrate from Africa to North. Could it be that memetic engineering induced this crucial step in evolution? Could it be that Neanderthals had to leave because they were not subject to this memetic engineering? Also the emergence of the first civilizations for about 10 thousand years ago might have involved memetic engineering. The ancient Sumerian myths about Gods who came from Heaven and made us their images might be memetic fossils reflecting what occurred.

6.5.5 Who performed the (memetic and) genetic engineering?

One can imagine two identifications for the ancient genetic/memetic engineers.

1. The guess that the engineers were extra-terrestrials (ETs) is supported by ancient myths. The Sumerian and Akkadian texts found inscribed on clay tablets, in which the role of the Elohim in Genesis is performed by the Anunnaki, tell about "Those Who From Heaven to Earth Came". These myths would relate to the last step in the sequence of engineering activities.

2. The second guess is that genetic engineering is due to a highly advanced civilization of a remote geometric future, perhaps futuro-terrestrials, and applying highly advanced technology based on time mirror mechanism and possibly utilizing simpler life forms, perhaps plasmoids, as their couriers. Abduction experiences might relate to genetic manipulations using plasmoids to do the hard job. In this case encounters with aliens would be based on sharing of mental images.

3. The third guess is that genetic engineering is self engineering. The work of Yu. Chen Kangeng gives evidence that the transfer of the genetic information by electromagnetic means is possible [J1]. According to [I12], where the method is summarized, the successful transfer of the genetic information from a donor bio-system to an acceptor system was achieved via high-frequency electromagnetic fields feed repeatedly through the optically-active donor bio-system and then delivered over a long period of time to the receiving bio-system in its early developmental stages. The hybrids created through the irradiation of eggs and seeds with such "genetically loaded" fields are claimed to show very specific mixed characteristics that were transferred to the next generation without need for further irradiation. This idea is discussed in [K25] on basis of a proposed realization of genetic code at the level of dark matter.

It would seem that the donor genome or parts of it are imprinted to the electromagnetic field pattern in the process and that this field pattern is able to modify the target genome.

Nothing precludes the possibility that genes/supergenes/hyper genes at some level of dark matter hierarchy can also code for genetic self engineering since these activities are after all very similar to other genetically coded bio-chemical activities. The computer analogy would be programs writing programs. The engineering genes would be activated by W MEs inducing plasma oscillation patterns. The claimed effects could be understood if the interaction with genetically imprinted electromagnetic field pattern activates genes inducing genetic self engineering yielding the genetic modifications consistent with the pattern represented by the em radiation.

Magnetic body would receive information about the desired outcome as electromagnetic field patterns emitted by other organisms, most naturally members of the same species. If these modifications are successful, the magnetic body is exposed to this information for long enough time to react and activate W MEs inducing the genetic program inducing the genetic program leading to the suggested genetic modification.

Hyper-genes integrating groups of organisms to larger wholes would be naturally involved with the mechanism. This mechanism would guarantee a rapid propagation of successful genetic modifications to the entire population and would be much more effective than the slowly occurring selection of random mutations. The possibly existing genes responsible for the genetic self engineering could be also introns and express themselves by activating nuclear RNA and process like reverse transcription.

REFERENCES

Mathematics

[A1] Redfern E Allenby R. *Number Theory with computing*. Edward Arnold, 1989.

[A2] Hofstadter DR. *Gödel, Escher, Bach: an Eternal Braid*. Penguin Books, 1980.

[A3] Bastin T et al. 7:445–, 1979.

[A4] Brillhart J et al. Factorizations of $b^m \pm 1$. *AMS*, 1990.

[A5] Noyes P. The Combinatorial Hierarchy - an approach to open evolution, 1980. Available at: `http://tinyurl.com/hszo9wb`.

Theoretical Physics

[B1] Waser A. The Global Scaling Theory: a Short Summary, 2004. Available at: `http://www.global-scaling.ch`.

Condensed Matter Physics

[D1] Stancil DD. *Theory of Magnetostatic Waves*. Springer Verlag, 1993.

Cosmology and Astro-Physics

[E1] Freeman PE et al. Examining the Effect of the Map-making Algorithm on Observed Power Asymmetry in WMAP Data. *Astrophys J*, 638, 2006. Available at: `http://arxiv.org/abs/astro-ph/0510406`.

Biology

[I1] Available at: `http://www.peter-hagemann.com/`.

[I2] Chromatin. Available at: `http://en.wikipedia.org/wiki/Chromatin`.

[I3] Cnidarians. Available at: `http://tolweb.org/tree?group=Cnidaria&contgroup=Animals`.

[I4] Identification and analysis of functional elements in one of the human genome by the ENCODE pilot project. Available at: `http://www.nature.com/nature/journal/v447/n7146/edsumm/e070614-01.html`.

[I5] Nucleosome. Available at: `http://en.wikipedia.org/wiki/Nucleosome`.

[I6] Supercoil. Available at: `http://en.wikipedia.org/wiki/Supercoil`.

[I7] The Genetic Code. Available at: `http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Codons.html`.

[I8] Dr. Phil Callahan on Power of Paramagnetism. *Nexus*, 2003. Available at: `http://www.nexusmagazine.com`.

[I9] Coghlan A. Junk DNA makes compulsive reading. *New Scientist*, 2608, 2007. . Available at: `http://www.newscientist.com/contents/issue/2608.html`.

[I10] International Human Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 2001.

[I11] Koruga D. Micro-tubule screw symmetry: packing of spheres as a latent bioinformation code. *Ann Ny Acad Sci*, 466, 1974.

[I12] Gariaev PP et al. Why are we still not able to successfully treat cancer and HIV?, 2001. Available at: `http://www.sciteclibrary.com/eng/catalog/pages/1171.html`.

[I13] Gariaev PP et al. The spectroscopy of bio-photons in non-local genetic regulation. *J Non-Locality and Remote Mental Interactions*, (3), 2002. Available at: `http://www.emergentmind.org/gariaevI3.htm`.

[I14] Kortschak RD et al. EST Analysis of the Cnidarian Acropora millepora Reveals Extensive Gene Loss and Rapid Sequence Divergence in the Model Invertebrates. *Current Biol*, pages 2190–2195, 2003. Available at: `http://tinyurl.com/py9buk9`.

[I15] Albrecht-Buehler G. Surface extensions of 3T3 cells towards distant infrared sources. *J Cell Biology*, 114:493–502, 1991.

[I16] Celera Genomics. *Science*, 291(5507), 2001.

[I17] Fröhlich H. The extraordinary dielectric properties of biological materials and the action of enzymes. *Nature*, 72(1968):641–649, 1975.

[I18] Sanduloviciu M Lozneanu E. Minimal-cell system created in laboratory by self-organization. *Chaos , Solitons & Fractals*, 18(2):335, 2003. See also *Plasma blobs hint at new form of life.* New Scientist vol. 179, No. 2413 p. 20 September 2003, page 16.

[I19] Volkenstein MV. *Biophysics.* Mir Publishers, Moscow, 1983.

[I20] Callahan P. *Insect behavior.* Acres U.S.A., 1971.

[I21] Gould SJ. *Wonderful Life.* Penguin Books, 1991.

[I22] Livshits V. Hemopoiesis in Figures and Tables. 2^n rule for Absolute Cell Number in Hemopoietic Organs. *Cell*, 1990.

Neuroscience and Consciousness

[J1] A method of changing biological object's hereditary signs and a device for biological information directed transfer. Patent N1828665. Application N3434801, invention priority as of 30.12.1981, registered 13.10.1992.

[J2] Pflanzenwachstum durch Elektrofeld. Available at: http://www.s-line.de/homepages/keppler/elektrofeld.htm. Instructions for the experiment yielding Ciba Geigy effect.

[J3] The Plants Respond: An Interview with Cleve Backster. Available at: http://www.derrickjensen.org/backster.html. Published in "The Sun" July 1997 and in "Free Spirit".

[J4] Selden G Becker RO. *The Body Electric: Electromagnetism and the Foundation of Life*. William Morrow & Company, Inc., New York, 1990.

[J5] Backster C. Evidence of a Primary Perception in Plant Life. *Int J Parapsych*, 10(4):329–348, 1968.

[J6] Blackman CF. *Effect of Electrical and Magnetic Fields on the Nervous System*, pages 331–355. Plenum, New York, 1994.

[J7] Martindale D. The body electric. *New Scientist*, (2447), 2004.

[J8] Nanopoulos DV. Theory of Brain function, Quantum Mechanics, and Superstrings, 1995. Available at: http://arxiv.org/abs/hep-ph/9505374.

[J9] Zipf GK. *Psycho-Biology of Languages*. MIT Press, 1965. Available at: http://linkage.rockefeller.edu/wli/zipf/. For references to Zipf's law see http://linkage.rockefeller.edu/wli/zipf/.

[J10] Narby J. *The Cosmic Serpent*. Jeremy P. Tarcher/Putnam, 1998.

[J11] Jaffe LF. Calcium Waves, 2001. Available at: http://waves.mbl.edu/calcium.waves.html.

[J12] Persinger M. The tectonic strain theory as an explanation for UFO phenomena, 1999. Available at: http://www.laurentian.ca/www/neurosci/tectonicedit.htm.

[J13] Skoulakis E Mershin A, Nanopoulos D. Quantum Brain ?, 2000. Available at: http://arxiv.org/abs/quant-ph/0007088.

[J14] Chomsky N. *Reflections on Language*. Pantheon Books, New York, 1975.

[J15] Westphal SP. Life still goes on without 'vital' DNA. *New Scientist*, (2450), 2004.

Books related to TGD

[K1] Pitkänen M. Crop Circles and Life at Parallel Space-Time Sheets. In *Magnetospheric Consciousness*. Available at: https:/tgdtheory.fi/pdfpool/crop1.pdf, 2006.

[K2] Pitkänen M. Crop Circles and Life at Parallel Space-Time Sheets. In *Magnetospheric Consciousness*. Available at: https:/tgdtheory.fi/pdfpool/crop2.pdf, 2006.

[K3] Pitkänen M. Did Tesla Discover the Mechanism Changing the Arrow of Time? In *TGD and Fringe Physics*. Available at: https:/tgdtheory.fi/pdfpool/tesla.pdf, 2006.

[K4] Pitkänen M. General Theory of Qualia. In *Bio-Systems as Conscious Holograms*. Available at: https:/tgdtheory.fi/pdfpool/qualia.pdf, 2006.

[K5] Pitkänen M. Genes and Memes. In *Genes and Memes: Part I*. Available at: https:/tgdtheory.fi/pdfpool/genememec.pdf, 2006.

[K6] Pitkänen M. Homeopathy in Many-Sheeted Space-Time. In *Bio-Systems as Conscious Holograms*. Available at: https:/tgdtheory.fi/pdfpool/homeoc.pdf, 2006.

[K7] Pitkänen M. Macroscopic Quantum Coherence and Quantum Metabolism as Different Sides of the Same Coin: Part I. In *Bio-Systems as Conscious Holograms*. Available at: https:/tgdtheory.fi/pdfpool/metab.pdf, 2006.

[K8] Pitkänen M. Many-Sheeted DNA. In *Genes and Memes: Part I*. Available at: https:/tgdtheory.fi/pdfpool/genecodec.pdf, 2006.

[K9] Pitkänen M. Massless states and particle massivation. In *p-Adic Physics*. Available at: https:/tgdtheory.fi/pdfpool/mless.pdf, 2006.

[K10] Pitkänen M. Matter, Mind, Quantum. In *TGD Inspired Theory of Consciousness*. Available at: https:/tgdtheory.fi/pdfpool/conscic.pdf, 2006.

[K11] Pitkänen M. p-Adic Physics as Physics of Cognition and Intention. In *TGD Inspired Theory of Consciousness*. Available at: https:/tgdtheory.fi/pdfpool/cognic.pdf, 2006.

[K12] Pitkänen M. Quantum Model for Hearing. In *TGD and EEG*. Available at: https:/tgdtheory.fi/pdfpool/hearing.pdf, 2006.

[K13] Pitkänen M. Quantum Model of EEG. In *TGD and EEG*. Available at: https:/tgdtheory.fi/pdfpool/eegII.pdf, 2006.

[K14] Pitkänen M. Time, Spacetime and Consciousness. In *Bio-Systems as Conscious Holograms*. Available at: https:/tgdtheory.fi/pdfpool/time.pdf, 2006.

[K15] Pitkänen M. Recent View about Kähler Geometry and Spin Structure of WCW . In *Quantum Physics as Infinite-Dimensional Geometry*. Available at: https:/tgdtheory.fi/pdfpool/wcwnew.pdf, 2014.

[K16] Pitkänen M. Topological Quantum Computation in TGD Universe. In *Genes and Memes: Part I*. Available at: https:/tgdtheory.fi/pdfpool/tqc.pdf, 2015.

[K17] Pitkänen M. Could Genetic Code Be Understood Number Theoretically? In *Genes and Memes: Part II*. Available at: https:/tgdtheory.fi/pdfpool/genenumber.pdf, 2019.

[K18] Pitkänen M. Does TGD Predict the Spectrum of Planck Constants? In *Hyper-finite Factors and Dark Matter Hierarchy: Part I*. Available at: https:/tgdtheory.fi/pdfpool/Planck.pdf, 2019.

[K19] Pitkänen M. Evolution in Many-Sheeted Space-Time. In *Genes and Memes: Part I*. Available at: https:/tgdtheory.fi/pdfpool/prebio.pdf, 2019.

[K20] Pitkänen M. From Principles to Diagrams. In *Towards M-Matrix: Part II*. Available at: https:/tgdtheory.fi/pdfpool/diagrams.pdf, 2019.

[K21] Pitkänen M. Nuclear String Hypothesis. In *Hyper-finite Factors and Dark Matter Hierarchy: Part II*. Available at: https:/tgdtheory.fi/pdfpool/nuclstring.pdf, 2019.

[K22] Pitkänen M. Quantum Hall effect and Hierarchy of Planck Constants. In *Hyper-finite Factors and Dark Matter Hierarchy: Part II*. Available at: https:/tgdtheory.fi/pdfpool/anyontgd.pdf, 2019.

[K23] Pitkänen M. TGD as a Generalized Number Theory: Infinite Primes. In *TGD as a Generalized Number Theory: Part I*. Available at: https:/tgdtheory.fi/pdfpool/visionc.pdf, 2019.

[K24] Pitkänen M. TGD as a Generalized Number Theory: Quaternions, Octonions, and their Hyper Counterparts. In *TGD as a Generalized Number Theory: Part I*. Available at: https:/tgdtheory.fi/pdfpool/visionb.pdf, 2019.

[K25] Pitkänen M. Three new physics realizations of the genetic code and the role of dark matter in bio-systems. In *Genes and Memes: Part II*. Available at: https:/tgdtheory.fi/pdfpool/dnatqccodes.pdf, 2019.