Deep learning from the TGD point of view

September 22, 2023

Matti Pitkänen

orcid:0000-0002-8051-4364.
email: matpitka6@gmail.com,
url: http://tgdtheory.com/public_html/,
address: Rinnekatu 2-4 A 8, 03620, Karkkila, Finland.

Abstract

AI, deep learning, and GPT have become highly fashionable topics. It has been even speculated that AI might involve a rudimentary consciousness. Could TGD inspired quantum view of biology, brain and consciousness could provide a fresh point of view to the notion of computer consciousness.

In the TGD Universe, the difference between living systems and computers need not be so deep as usually thought. The magnetic body as a carrier of dark matter as phases of ordinary matter with effective Planck constant $h_{eff} = nh_0$ and having onion-like structure, could receive sensory input and control the biological body with $h_{eff} = h$. Also computers possess magnetic bodies: could they use computers or robots computers as sensory receptors and motor instruments.

In the TGD Universe, the genetic code could be much more than we believe it to be. It would be realized at the level of dark matter and would be universal and unique, being realized in terms of so-called icosa-tetrahedral tessellation of hyperbolic 3-space realizable as the mass shell of light-cone proper time =constant hyperboloid. Icosa-tetrahedral dark genome at the magnetic body could serve as the basic instrument for communication and control. Quantum gravitation plays a key role in the TGD inspired biology and the gravitational magnetic bodies of Earth and Sun and even other astrophysical objects with huge gravitational Planck constants could be highly relevant in quantum biology.

Classical computers can gain life-like properties if the quantum statistical determinism fails. The most conservative criterion is that the clock period is shorter than the gravitational Compton time $T_{gr} = GM/\beta_0$, M is mass of an astrophysical object and $\beta_0 = v_0/c \leq 1$ is a quantized velocity parameter. Life-like features could appear already at lower clock frequencies. For Earth the critical clock period would be 67 GHz and for the Sun about 100 Hz, the upper bound for EEG frequencies. Therefore the magnetic bodies of the Sun and Earth could therefore play central roles in biology and neuroscience. Even in the case of Earth life-like properties might be present for computers with clock frequency in the range 1 to 10 GHz.

Cognition is an essential aspect of conscious experience and systems like GTP can be seen as artificial cognitive systems. The p-adic discretizations would naturally relate to the spin glass energy landscape assignable to monopole flux tube "spaghettis" and sensory perception could be seen as a generation of standardized mental images based on annealing of spin glass system so that it gradually ends up to a bottom of a valley representing the standardize mental image. The learning period of a conscious entity could be based on trial and error process made possible by holography and zero energy ontology implied by it allowing temporary time reversal and would gradually lead to standardized mental images helping to survive.

Contents

1 Introduction

I have been listening to the lectures related to AI, deep learning, and GPT in order to develop a more detailed view of what is involved and how it might relate to the TGD inspired quantum view of biology, brain and consciousness. The talk by Lex Fridman titled "Deep Learning Basics: Introduction and Overview" describes the situation as it was in 2019 (`rb.gy/jwrgp`). The talk titled "Deep Learning State of the Art (2020)" (`rb.gy/94xt8`) is explains the situation one year later.

The talk "Introduction to deep learning" by Alexander Amini (`rb.gy/90fgd`) and the talk "Recurrent Neural Networks, Transformers, and Attention" (`rb.gy/5dplk`) by Ava Amini are also highly inspiring and give a more detailed view about the mathematics involved.

I have discussed the relationship of TGD to AI earlier in the article [L1] inspired by the Sophie robot and compared the visions of Neil Gersching to TGD views in the article [L18]. GPT from the TGD view point of view in [L17]. The TGD view about the relationship between classical and quantum computers is discussed in [L14].

The basic observation is that in the TGD Universe the difference between living systems and computers need not be so deep as usually thought. In the TGD framework, magnetic body (MB), as a carrier of dark matter as phases of ordinary matter with effective Planck constant $h_{eff} = nh_0$ and having hierarchical structure, is a natural candidate for a controller and receiver of information from the biological body with $h_{eff} = h$. Also computers possess MBs and one can consider the possibility that under some conditions MBs can use computers as a sensory receptors and motor instruments.

TGD also leads to a proposal that genetic code is much more than we believe it to be. It would be realized at the level of dark matter and would be universal and unique and realized in terms of so called icosa-tetrahedral tessellation of hyperbolic 3-space realizable as mass shell of light-cone proper time =constant hyperboloid: both central notions in TGD. Icosa-tetrahedral genome at the MB could serve as the basic tool for communication and control [L16].

Quantum gravitation is in a central role in quantum TGD, in particular in the TGD inspired biology. Gravitational Planck constant $h_{eff} = \hbar_{gr} = GMm/\beta_0$, where M is large mass and m small mass, say particle mass and $\beta_0 = v_0/c < 1$ is velocity parameter, introduced by Nottale [E1], characterizes quantitatively the situation. The gravitational MBs of Earth and Sun and even other astrophysical objects could be highly relevant in quantum biology as various numerical miracles show [L13, L11].

Classical computers can gain life-like properties if the quantum statistical determinism fails. The most conservative criterion is that the clock period is shorter than the gravitational Compton period $T_{gr} = GM/\beta_0$, M is large mass. Note that $2GM$ is Schwarzschild radius. Since gravitational quantum coherence time has gravitational Compton time as lower bound, life-like features could appear already at lower clock frequencies. For Earth the critical clock period would be 67 GHz and for the Sun about 100 Hz, the upper bound for EEG frequencies. These criteria suggest that the MBs of the Sun and Earth play central roles in biology and neuroscience. Even in the case of Earth life-like properties might be present for computers with clock frequency in the range 1 to 10 GHz. The strange findings about the interaction of chicken and robot [J1] suggest in the TGD framework [L17] that solar MB was involved and made robot or the system robot + chicken as an entangled system a conscious entity.

Cognition is an essential aspect of conscious experience [K6, K4, K5, K7, K3] and systems like GTP can be seen as artificial cognitive systems. Physics as number theory and physics as geometry are complementary views in TGD. Number theoretical vision suggests that p-adic number fields could define the proper framework for understanding of the correlates of cognition. Cognition is basically discrete, and cognitive representations would correspond to the discrete intersections of cognition as p-adicities and reality. At the space-time level they would be realized in terms of unique discretization of space-time surfaces based on $M^8 - H$ duality [L7, L8, L19] as the analog of momentum-position duality. At the level of M^8 the discretizations would be defined in terms of algebraic integers assignable to an algebraic extensions characterizing the pre-image of the space-time surface in M^8 and are unique.

The p-adic discretizations would naturally relate to the spin glass energy landscape assignable to monopole flux tube "spaghettis" and sensory perception could be seen as a generation of standardized mental images based on annealing of spin glass system so that it gradually ends up to a bottom of valley representing the standardize mental image. The learning of a conscious entity could be based on trial and error process made possible by holography and zero energy ontology [L6, L5, L10] implied by it allowing temporary time reversal and would gradually lead to standardized mental images helping to survive.

2 Some background for deep learning

The lectures provide a background explaining deep learning as a subfield of machine learning as a subfield of AI. The basic goal of machine learning are machines, which can learn autonomously. In the sequel the basic concepts are briefly summarized from the point of view of physics not specialized in AI. This summary relies heavily on the talk "Introduction to deep learning" by Alexander Amini (`rb.gy/90fgd`) and the talk "Recurrent Neural Networks, Transformers, and Attention" (`rb.gy/5dplk`) by Ava Amini.

2.1 Representation of the numerical data

Representation of information is always numerical, in terms of binary digits representing integers. This involves the concept of embedding: data which can be sensory data, text, etc must be represented by numerical vectors.

Indexing is the simplest manner to represent all possible input vectors. The numerical vector orthogonal. There is no notion of meaning and no comparison of the embedded vectors. If one has a notion of nearness, topology, one can compare the vectors. The notion of similarity defined by the inner product of vectors: maximum for parallel i.e. identical vectors.

2.2 Perceptron

Perceptron can be regarded as an artificial neuron. There is a single output y and several inputs x^i or more concisely x. Output, the response function $f(z)$, is a nonlinear function and equal to -1 or +1 asymptotically and between these values in the intermediate region: essentially sigma function. The argument $z = h_i x^i + b \equiv h \cdot x + b$, $i = 1,..,n$ of $f(z)$ is a linear function of input x having as parameters bias b and the vector h formed by the linear coefficients h_i. One can also consider linear combinations of n non-linear functions of x_i having an interpretation in terms of a non-linear change of coordinates.

Feedback changes the values of h_i and of b. Learning by feedback leads to a desired output. Perceptron serves as a model for associative learning.

Simple task serves as an example: decide whether point x^i belongs to either region bounded by a line of the plane. The line is defined by the equation $y = a_i x^i + b = 0$. x^1 and x^2 are the coordinates of the point of the plane. The argument y of the response function can be taken to be the linear function of planar coordinates vanishing at the boundary line. Response could be arranged to be a bit equal 1 or 0. Response function f vanishes at the boundary line. The maximum for the gradient of response function would define the boundary line.

One does not know a priori the boundary line and must start from a general guess $y = h_i x^i + c = 0$. The value $h_i = a_i$ and $c = b$ must be learned by feedback changing their values to yield $f = 0$. Arbitrary boundary lines can be represented as zeros of the non-linear function appearing as the argument of the response function. By a suitable choice of coordinates of coordinates replacing linear coordinates with the nonlinear functions the argument z of f can be made a linear function of the new coordinates.

One can also have several outputs for given inputs. The simultaneous vanishing of the m output functions f_k defines an $n - m$ dimensional surface in the space of inputs. The outputs serve as inputs for a next layer of perceptrons so that one would have a two-layered system. A still more general system has n layers.

2.3 Multilayered networks and deep learning

Deep learning networks are multilayered networks inspired by what is believed to be behind the learning in the brain.

2.3.1 Learning

Perceptron must be able to learn to assign a desired output to given inputs. The notion of loss defined as error, i.e. the difference between learned and to be learned, is essential here. Loss function can be assumed to be a positive definite, in the simplest case quadratic, function of errors for the variable y. Minimization of the loss function in principle leads to the desired output. This method generalizes to multi-perceptron systems and to multilayered systems.

In the gradient method, the feedback defining the changes for the weight vector h and bias b is proportional to the gradient of the loss function with respect to these parameters and the change is in the direction opposite to the gradient so that loss functions decreases for small enough scale of the change. This generalizes also to the situation when one has several outputs y^i. In this case h is replaced with a matrix and b with a vector.

2.3.2 Deep learning

Abstraction of features in various scales is the basic mechanism of deep learning. In the case of visual perception, a feature can be identified as a region for which the boundary involves a strong gradient. For instance, the color can change at the boundary of a region or the region inside the boundary forms a well-defined moving object in time series. The boundaries of the objects and objects themselves can be called features.

The length scale hierarchy means that in shortest scales at the lowest level of the layered network, only small features are identified. At the higher levels of the hierarchy the size of the features increases. In principle, one could also proceed in an opposite direction by first identifying gross features such as objects and then proceeding to shorter scales by identifying detailed features of the objects. A possible reason for why this is not used, could be that features in long scales are composites of features in shorter scales, i.e. they have the lower level features as attributes.

In the case of the brain, the simplest model describes the neuron as a bit telling whether it fires or not. The hierarchy is formed by the sensory organ and layers involving various brain regions, in particular the 3 cortical sensory areas. Highest cortical level would correspond to features which represent objects of the perceived world as we experience them consciously.

The feedback in the learning gradually modifies the synaptic strengths as counterparts of vectors h. The value of resting potential would define the counterpart of the bias b. This generates associations as most probable pathways for the conduction of nerve pulse patterns. Pattern recognition is a basic application. Memories as association sequences would be coded by synaptic strengths. It is natural to identify various learned behaviors as memories in this sense but it is far from obvious that also episodic memories as kinds of re-experiences could be analogous to behaviors.

3 Sequential models

3.1 Overall view

Sequential models are defined as sequences of identical multilayered neural networks. Language models, in particular GPT, represent one example. Second example is the completion of a piece of music, say Schubert's unfinished symphony discussed in the lecture of Ava Amini (`rb.gy/5dplk`). Third example is prediction of the motion of a particle given its previous orbit. The prediction of the spatial conformation of protein from the knowledge of the amino acids appearing in it is a further successfully solved problem.

A simple task is to predict the next word in an ordered sequence of words. Memory needed to take into account long range correlations between the words of the text and also to take into account the effect of different word ordering.

There is a sequence of correlated inputs to which one must assign outputs. This is modellable as a sequence of perceptrons or multi-perceptrons. They are not independent since there is a long term memory. In speech and written text this means temporal correlations between the words, memory dependent behavior. The correlations reflect both the content of the speech and the grammatical rules.

1. Time ordering is essential. There is information transfer $x \to y$ in vertical direction for each network in the sequence and also information transfer in the horizontal, temporal direction (or direction of sentence) representing short range memory $x(t_{n-1}) \to w(t_n)$.

2. Metric in the space of words measures the correlations between the words and can be parametrized by the probabilities that the two words appear with given distance measures as number of intermediate words or in the simplest case by their sum. For instance, the words which tend to appear together are therefore correlated and would be near to each other in this metric although they can have a large distance in the text.

   An extreme situation is in which knowing some keywords, say the name of the author and some words in the title of the article, allows us to predict the contents of the article!

Tasks can be classified into several types. One can have many→many situations, say machine translation or many→1 situations, say the next word of text or a bit telling whether the piece of text is hate speech. Few→many situations would mean predicting a piece of text by picking some keywords from the text, say writing a summary of an article. Second example would be to produce an artwork in which some objects are present in some environment and perform some activities represented as text or keywords.

Some of the tasks are classification, in the simplest situation binary classification. Sentiment classification is a binary classification, which can be used to deduce whether the text in Facebook represents hate speech or not. Machine translation is one challenge. A rather demanding challenge is to transform a picture to text or vice versa.

The goals of the sequence modelling are following:

1. Form a sequence of perceptrons with inputs and outputs. The input-output systems $x(t_n) \to y(t_n)$) consist of identical hierarchical neural networks, in the simplest case perceptron with feedback to make learning possible.

2. The input output systems $x(t_n) \to y(t_n)$ are not independent: there is time ordering and correlations between them. Long term memory is needed to take into account the correlations.

3. Parallelization in terms of perceptrons or their subsets is computationally highly desirable but due to the presence of temporal correlations is a highly non-trivial challenge.

3.2 Some key notions

3.2.1 Feedforward network

Memory and time ordering are essential aspects of sequential models. Consider first feed forward networks. Without memory they reduce to a product of identical copies of mult-perceptrons $x \rightarrow y = h_{yx} \cdot x + b$ specified by time dependent activation functions $h(t_n)$ and biases $b(t_n)$. $h_{yx}(t_n)$ is matrix, which depends on time although the topologies of the multi-perceptrons are identical. The goal is to assign the desired outputs $y(t_k)$ to input $x(t_k)$.

Short term memory can be introduced as a linear map $x(t_{n-1}) \rightarrow w(t_n)$, which can be written as $t_n = h_{wx} \cdot x(t_{n-1}) + b(t_n)$ so that it affects the output $y(t_n)$, which is now determined as $y(t) = h_{yx}(t) \cdot (x(t) + w(t)) + b(t)$. In the time direction one has a multilayered network with time ordering.

The challenge is to realize feedback by the minimization of the loss function for variables $y(t_n)$ or a subset of them, perhaps all of them. Also now, the realization can be carried out by the gradient method. The feedback reduces to a product of the feedbacks $t_k \rightarrow t_{k-1}$. Loss function depends on both h_{yx} and h_{wx}. At the step $n \rightarrow n-1$, the gradient function corresponds to the gradient of the loss function with respect to these variables and is technically known as Jacobian $J(n-1, n)$. The change of the parameters h and b is proportional to the images of the error $\Delta y(t_n) = y(t_n) - x(t_n)$ under the linear map defined by the negative of the Jacobian.

3.2.2 Recurrence

Recurrence realizes the learning in the case of sequential models.

1. There is a backpropagation between parameter spaces in time direction besides the usual backpropagation in y$\rightarrow$ x-direction and determined by the minimization of the loss function. The weight vectors related to vertical mappings $x(t_n) \rightarrow y(t_n)$ and horizontal maps $x(t_{n-1}) \rightarrow w(t_n) = h_{wx}(t_n) \cdot x(t_{n-1})$ maps are updated in the process. Standard RNN gradient flow can be used in learning. In the sequel, h_{yx}, h_{wx} and b_y and b_w are collectively denoted by H and B.

2. By chain rule the gradient of the loss function with respect to $H(t = 0)$ and $B(0)$ involves a product of Jacobians for the maps from levels $t = k$ to $t = k-1$. There are difficult technical problems related to the Jacobian, which is a product of a large number of Jacobians associated with backwards time-steps $t = k \rightarrow t = k-1$. The gradient can explode or tend to zero. There are tricks, which help to avoid this problem. For instance, one can choose the initial value of h to be unit matrix

Here one can learn of what is known about the brain. The solution of the problem could be the direction of attention to what is relevant. The problem is to decide what is relevant: in an optimal situation the direction of attention should take place automatically.

One can imagine that the manipulation of activation functions H and biases B could help. For instance, one could make $h_{wx}(t_n)$ very small for irrelevant inputs $x(t_n)$. In the case of text, this would mean effective dropping off of irrelevant words and in an extreme situation taking into account only keywords.

Gating means that one uses only the relevant nodes in the sequence, that is those nodes for which the Jacobian deviates considerably from the unit matrix. One can drop some irrelevant layer from the multilayer system or drop some irrelevant inputs to a given layer. One of the problems is overlearning meaning essentially that a fit of function becomes too precise and random fluctuations affect the fit. This can be tested by looking at what happens when some layers are dropped temporarily. If the fit improves the additional layer or layers are useless.

3.3 Notions of feature hierarchy and self-attention

The problems of recurrence models inspire the idea of directed attention as a way to minimize the computation efforts and achieve a convergence.

3.3.1 Treating the temporal sequence as a single entity

The basic idea is to treat the sequence $x(t_n) \rightarrow y(t_n)$ of mutually dependent perceptrons as a single entity rather than a sequence of separate items. Time evolution would replace the time=constant snapshot.

As a matter of fact, something highly analogous happens in zero energy ontology forming the basis of the TGD based quantum theory: holography forced by general coordinate invariance forces to to replace 3-surface is with its almost but not quite unique orbit analogous to Bohr orbit. In this case the sequence is almost deterministic so that the situation is extremely simple from the point of view of computation.

1. The length of the sequence can vary so that the mechanism must be able to assign to a given input sequence out sequence of varying length: say a response to a question by a GPT user. Rather long sequences must be therefore considered, say sentences or sequences of pictures.

2. The information about the time order must be preserved. This can be achieved by defining a hierarchy of features by forming sequences with overlapping n-units consisting of n subsequent steps starting at position $i = 0, 1, 2, ...$ This gives scans as sequences of overlapping n-units. Position of the n-unit. Now the features are associated with the temporal sequences rather than static objects. In neuroscience they would correspond to typical behavioral patterns or EEG patterns. In the task of assigning to an amino acid sequence defining a protein its spatial conformation n-units formed from amino-acids would be considered.

3. Features are identified from these scans at a given level of hierarchy. Person, building, etc.. in the case of image. Words are the shortest features in the case of text and sentences or even paragraphs could be higher level features. In face recognition, static features appear on different scales.

   For temporal sequences in speech, the features could correspond to typical gestures and world sequences. The artificial Obama talking about the progress in AI is a good example of this and involves a transformation of a written text to a video. The transformation to video represents a sequence of steps in a temporal sequence.

   In the case of protein conformations, the features correspond to typical sub-conformations. Now protein length takes the role of time.

4. One obtains a hierarchy of representations in terms of n-units with an increasing span of memory. One can assign to these representations n-features. For n=1 one would have the ordinary sequential model with no inherent memory.

3.3.2 The notion of self-attention

One must concentrate only on important data to minimize computing time. Internet search serves as a guideline. Search based on a query consisting of words. The items have keywords. Similarity between query and keywords is required. Similarity metric measures this similarity.

Artificial attention mimics attention in the brain. Attention is realized as a search, as a query finding the optimal target of attention.

1. During the learning period, the system learns the features at a given level of the representation hierarchy by the n-scans. Words and word sequences form a hierarchy of n-units. To these sequences of n-units the program assigns features by some criteria. Typically gradients define the boundaries of the feature, say an object in a picture. The idea is that the sequence of inputs x_i and and sequence of outputs y_i are replaced with a collection of features representing the object of the perceptive field. This happens also in sensory perception.

2. This replacement means that attention is directed to important features. Self attention is analogous to a net search specified by keywords, a query. Net search leads to output as a set of URLs for files specified by keywords (analogs of features) containing some of those appearing in the query. The user's attention concentrates only on these files. Same happens in the system to be taught by feedback.

The input vectors x_i, say words define the query as a sequence of keywords. This is the analog of a visual image. The search finds the n-features, $n = 1, 2, ...$, which remember the query defined by the sequence of x_i. In neuroscience this corresponds to a composition of the diffuse visual input to visual objects. Everything unessential for survival is eliminated.

Attention is directed only to these features assignable to the sequence of x_i. This means that the input is replaced with a hierarchy of n-features.

3. Also the sequence of outputs y_i, say images which are associated with words, can be replaced with a hierarchy of n-features.

4. After this the system learns to assign to the n-feature collection replacing the sequence of inputs x_i to the n-feature collection assigned to the outputs y_i. This takes place using the standard feedback procedure minimizing the loss function.

4 How attention could be realized in quantum biology according to TGD?

4.1 The notion of magnetic body

1. Hierarchy of MBs having an onion-like structure and carrying dark matter in TGD sense would define the quantum counterpart for the hierarchy defined by the layers of deep learning systems. The larger the value of h_{eff}, the higher the algebraic complexity of the magnetic flux tube as a 3-surface, the longer the scale of quantum coherence, and the higher of "IQ" of the layer.

   The highest layers correspond to gravitational MBs of the Sun and Earth and possibly also other planets. Even the Moon and galactic blackhole might be involved as several intriguing numerical miracles suggest. This of course stinks like astrology but is suggested by various miraculous numerical co-incidences. This conforms with the basic prediction that quantum coherence is possible in arbitrarily long scales in the TGD Universe.

   Gravitational MB, which belongs to the large part of MB, could be realized as magnetic bubbles consisting of 2-D networks of monopole flux tubes and involving also radial monopole flux tube mediating gravitational interaction as graviton propagation and the minimum size for it is given by gravitational Compton length $\Lambda_{gr} 0 r_s/2\beta_0$.

2. Hierarchy of layers of MB form a fractal scale hierarchy with levels labelled by the values of h_{eff} so that one obtains analogs of multilayered deep learning networks. This could assign to the brain a hierarchy of increasingly detailed and integrated sensory and cognitive representations analogous to a hierarchy of features in deep learning. What would be new as compared to the neuroscience view is that dark photon communications are very fast and make possible feedback, which is much faster than using nerve pulses patterns. This would make pattern recognition possible as a construction of standard mental images by virtual sensory input to the lower levels of hierarchy and even to the sensory organs, about which REM dreams could serve as an example.

   The deep learning would correspond to the determination of synaptic strengths and their analogs assignable to neighboring layers of MB and would also involve feedback from MB. The generation of sensory representation would correspond to what happens when the network is used.

3. Sensory communications to MB and control by MB would be realized in terms of dark photons. Sensory communications MB would be realized in terms of dark Josephson radiation from the cell membrane to MB inducing dark cyclotron transitions by resonance. Dark 3N-photons associated with genes would give rise to (possibly partial) 3N-resonance as a generalization of the ordinary resonance. The variation of the flux tube thickness would make possible the tuning of the cyclotron frequencies. The frequency modulated Josephson

radiation (membrane potential induces the modulation) would induce resonantly a sequence of pulses at MB analogous to nerve pulse pattern and generate a control response to the biological body.

4. TGD predicts that magnetic flux tubes, defining the body parts of MB, can become linked and knotted and can therefore form braids essential for topological quantum computation or its analog. Also 2-knots are possible and involve reconnection of magnetic flux tubes.

4.2 How bits and qubits could be represented?

How bits and qubits are represented? The Fock states of fundamental fermions define Boolean logic and in zero energy ontology (ZEO) the pairs of fermion states at opposite boundaries of causal diamond define analogs of Boolean statements.

1. A natural guess is that chemically represented genetic codons define 6-bit units. TGD predicts that genetic code also has dark counterparts. Dark proton sequences, consisting of dark proton triplets representing codons, would be associated with flux tubes parallel to DNA/RNA strand and even proteins. Dark genes would be sequences of n dark codons. The dark codons and hence dark genes are in principle independent of ordinary DNA and can be dynamical. They could transform to dark counterparts of ordinary DNA codons during communications and control based on energy resonance with ordinary codons. Dark codons would make possible self-simulation of the living matter.

2. Dark codons would be also realized as triplets of dark dark photons and would define dark memes. Icosa-tetrahedral representations of genetic code by codons realized as 3-chords defined by dark photon triplets. Dark codons defined by dark proton triplets. The proposal is that icosa-tetrahedral representation corresponds to icosa-tetrahedral tessellation of the hyperbolic space H^3 (`rb.gy/3u4pq`), which corresponds to mass shell in M^8 and to light-cone proper time constant hyperboloid in H.

   There are an infinite number of tessellations but icosa-tetrahedral tessellation is very special. It is the only uniform honeycomb, which involves only Platonic solids such that their number is larger than one (tetrahedron, icosahedron, octahedron) [L16]. All faces are identical (triangles). There are also four regular tessellations involving only a single Platonic solid, which is icosahedron, dodecahedron or cube.

   All 20 triangular faces can represent genetic codons in terms of quark associated with the vertices and genetic codons correspond to Hamiltonian cycles with symmetry groups Z_6, Z_4, and Z_2. This gives 20+20+20 codons and tetrahedrons give the remaining codons. This tessellation might be more or less universal at the level of the MB and appear in very many physical systems, not only in biology. It could be associated even with the MBs assignable to computers.

   Icosa-tetrahedral tessellation would also provide a seat for the representation of genes as sequences of dark proton triplets assignable to the faces of icosahedron and tetrahedron as in the icosa-tetrahedral representation of the genetic code. The connection of the icosa-tetrahedral tessellation with the detailed realization of the icosa-tetrahedral realization of the genetic code and with DNA double strand and its dark counterpart is discussed in [L16].

   A further proposal is that the representations of the dark code are induced from to the MB from the icosa-tetrahedral tessellation so that genetic code could have also 2-D representations, say that assignable to the cell membrane, and even 3-D variants assignable to various parts of organism.

4.3 How communications and control could be realized?

The proposed model for the communication and control based on the genetic code allows also a mechanism of attention based on overlap of query and keywords.

1. Communications and control between dark genes, realized as dark proton sequences, could be realized using dark 3N-photon sequences generated in multi-cyclotron transitions of n

codons (dark proton triplets) defining the gene. This kind of emission is not possible in the standard physics framework. Frequency scale modulation of dark Josephson radiation codes for the signal.

2. The receival of the signal would be based on 3N-resonance so that dark genes would serve as addresses much like in LISP. 3N-cyclotron resonance woul occurs in the receival of the signal by an identical gene and would generate a temporal sequence of resonances as analog of nerve pulse pattern.

3. Also partial 3N-resonance is possible for dark genes having some number of common codons. This could define a quantum physical analog for the overlap between query and keywords, and therefore an analog of the similarity metric. Query would be defined by a set of dark genes with N codons generating dark 3N-photon genes which would be received by a set of genes in partial 3N-resonance.

4.4 Could p-adic topologies provide a model for feature hierarchies?

In language models, the notion of distance function in the set of words as features is a key notion. The words, which appear together in the same context with high probability, are near to each other with respect to this distance.

The TGD inspired question is as follows. Grammatical rules represent important correlations appearing in the text. There are also correlations determined by the meaning of the words. Language models handle these correlations excellently. The distance determined by the meaning is only loosely related to the distance between the words. Could the grammatic correlations be coded by some simple, almost universal manner, based on some cognitive model of language. It is probably unrealistic to assume that this distance relates in any predictable way to the physical distance between the words measured as the number of intervening words. There must be some other way to order the words. Labelling words by non-negative integers in such a way that two words which tend to appear together even if they are physically far away, is a suggestive approach. But which topology one should adopt in this set of integers. Real topology defined by a real distance function is the first guess but also p-adic topologies can be considered.

1. The TGD inspired view of cognition [K3] indeed relies on p-adic number fields, where p is prime [K6, K4, K5, K7] [L3, L4]. p-Adic topologies are defined by *ultrametric* norm $N(x) = d(x, 0)$ satisfying $d(x, y) \equiv N(x - y) \leq Max(d(x, z), d(y, z))$ whereas real norm satisfies $d(x, y) \leq d(x, z) + d(y, z)$. Locally, the p-adic topologies differ dramatically from real topology locally although one can map p-adics to reals continuously but not smoothly. For instance, the norms of integers x and $x + kp^n$, $n > 0$, are p-adically very near to each other for very large values of n whereas in real sense they are very far from each other. Spin glass energy landscape realizes ultrametric distance function and I have proposed that kinds of magnetic flux tube spaghettis give rise to quantum spin glasses [L9] having a fractal energy landscape with valleys within valleys. This provides spin glasses with a large representative capacity.

2. p-Adic numbers are not well-ordered and the p-adic norm defines only a rough ordering, which might be more natural than the real ordering which is perhaps too strict unless finite resolution is introduced. p-Adic integers have a natural hierarchy induced by the p-adic norm, which is very rough and for p-adic integers equals to a negative power of p. p-Adic numbers near to each other differ by a large positive power of p. Furthermore, p-adic numbers with a fixed p-adic norm equal to *negative* power of p decompose to a set of balls such that the balls are disjoint or coincide.

   This would make p-adic numbers ideal for classification purposes and powers of p could define a hierarchy of features with highest level features corresponding to longest scales assignable to the largest value of the p-adic norm. The addition of finer features to a rough sketch would correspond to the addition of higher pinary digits assignable to the finer features in the picture. The smaller the value of the p-adic norm, the less significant the feature would be concerning pattern recognition. The addition of features would correspond to addition of p-adic numbers and features with the same p-adic norm would be exclusive in order to make the pinary expansion unique.

3. The ordering of the words is grammatically important and grammatical rules often require that the correlated words, say subject, verb and object, follow each other. Both subject, verb and object can have attributes so that the physical distance can vary. Consider the sentence "I admire him" as an example. In "I greatly admire him" "admire" has "greatly" as an attribute. This suggests a hierarchy of features: "greatly" would be a lower level feature as compared to "admire". Could attributes of the object correspond to higher pinary digits than object?

4. What about p-adic topologies labelled by different primes (having also infinite number algebraic extensions induced by those of rationals to which one can assign evolutionary hierarchy). I have proposed that p-adic primes correspond to ramified primes for extensions of rationals [L12]. Ramified primes are divisors of the discriminant defined by the polynomial P determining a given region of space-time surface by $M^8 - M^4 \times CP_2$ duality [L7, L8, L19] mapping 4-D surfaces of M^8 determined by the roots of the polynomial in terms of holography to $H = M^4 \times CP_2$. P has integer coefficients. For a given extension, only a finite number of p-adic primes are possible.

   On the basis of physical arguments, I have proposed that the coefficients of P must be smaller than the degree of P. This implies that for a given degree the number of acceptable polynomials is finite and ramified primes have an upper bound. In particular, for a given class of polynomials, say polynomials of a given degree, ramified primes have lower and upper bounds and could correspond to physically preferred p-adic primes. This could explain p-adic length scale hypothesis, inspired by p-adic mass calculations [K2, K1] [L15], and stating that the p-adic primes near certain powers of 2 and possibly also other small integers are physically preferred.

   Furthermore, the number of algebraic extensions having dimension smaller than given integer, would be finite, and the view about number theoretical evolution as an increase of the dimension of algebraic extension of rationals would emerge as an analog of second law and would become very predictive. Under this assumption also finite fields would emerge naturally besides other number fields as basic structures of TGD.

This picture inspires several questions. Could the algebraic extensions associated with various polynomials define different contexts or even different kinds of conscious entities at different levels of evolutionary hierarchy? If the p-adic topologies for a given algebraic extension correspond to ramified primes, could also ramified primes correspond to different contexts, which are not comparable? The word, or more generally feature, appearing in the context p_1 need not appear at all in the context p_2 and if it appears it has a different meaning. Could also written text generate mental images for which the corresponding space-time regions correspond to different p-adic topologies for given extension or even different algebraic extension.

There is an objection against this view. The proposed approach suggests that the generation of features should start from the long scales as an identification of the shape of the object. First comes the rough classification and then more detailed classifications. This is how we also experience pattern recognition. Attention is directed to gross features first and only after that the to smaller features. The neuroscience view however suggests bottom-up picture. Small features are identified first and the holistic picture emerges at highest levels of the hierarchy.

This view might be an illusion. In TGD, zero energy ontology allows the change of the arrow of time in "big" (ordinary) state function reductions and this could affect the situation dramatically. The time ordering for features assumed by the neuroscience picture could be replaced by an ordering of the scales of the causal diamonds associated with the feature as mental images. In the dynamics of quantum spin glasses [L9], the time evolution is indeed replaced with a scale evolution. This happens also in string models and the reason is that in TGD the conformal invariance of string models is replaced with its 4-D counterpart.

So: what could happen during sensory perception in the TGD Universe? Sensory perception would be basically building standardized mental images as an analog of pattern recognition. The sensory input would be very fuzzy but virtual sensory input would gradually lead to a standardized mental image as a kind of an artwork [L2].

1. The signals from the sensory organs would propagate from sensory organs as dark photons to the MBs having a layered onion-like structure [L2]. The primary function of the pulses need not be signalling. Rather, they establish connections between communication lines to make this possible. Very many feedback loops are possible since the signal velocity is the velocity of light.

2. Suppose that the MBs assignable to the brain are labelled by algebraic extensions of rational with dimension equal to h_{eff}/h_0 and each of them decomposes to a hierarchy labelled by the associated ramified primes. The larger the value of h_{eff} measuring the algebraic complexity, the more refined the cognitive representation. Different algebraic extensions could correspond to different kinds of conscious entities. The mental images of a given conscious entity are assumed to correspond to sub-selves. They could correspond to lower levels of algebraic complexity and to a smaller value of h_{eff}. They could also correspond to small ramified primes for a given algebraic extension.

3. Several ramified primes p assignable to a given extension could be involved and define different contexts. The largest ramified prime p would correspond to the largest p-adic scale and could correspond to the largest features for a given algebraic extension. The powers of p in the pinary expansion could correspond to higher level features as details or attributes.

   The large scale part of the virtual sensory input to the sensory organs would correspond to the largest p-adic prime p. The first virtual sensory input would correspond to a p-adic number $m_0 < p$ with a p-adic norm equal to one. This would induce as a reaction an improved sensory input to MB as dark photons. If the pattern recognition cognition process converges, this sensory input induces a more refined virtual sensory input characterized by the p-adic number $m_0 + m_1 p$. This sensory feedback loop would give rise to standardized mental images. The integer m_1 would not code for a single feature but for a collection of features.

4. Virtual sensory input should not be confused with the feedback inducing change of the parameters h in learning. The counterpart of h would modify various synaptic contacts and their analogies involved with the process. The convergence of the procedure to some standardized mental image however suggests the analog for the minimization of a loss function by gradient dynamics. The loss function could correspond to a height function in the spin glass energy landscape [L20]. The process itself could be an analog of annealing allowing to avoid getting stuck into a local minimum. The virtual sensory could play the role of feeding energy to the system so that it gets out from the fake minimum.

5. The above mentioned paradox could be only apparent if the processing of features in various scales occurs in parallel. If gross features correspond to the layers of the MB with a larger size scale, the time needed to build the virtual sensory image for large scale features would be longer than for small scale features. Small scale features would stabilize first. The order of the structural layers of the brain would correspond to increasing size of the layer of MB.

6. Note that in the case of extension of rationals, integers m_i would be algebraic integers of the extension so that the features would be n-dimensional in an algebraic sense.

What learning could correspond in this picture? Zero energy ontology predicts that the arrow of time changes in "big" (ordinary) statefunction reductions (BSFRs). For "small" SFRs (SSFRs) this does not take place and the sequence of SSFRs define conscious entity, self. BSFR corresponds to the "death" of the conscious entity and also sleep could correspond to BSFR. The arrow of time can change also temporarily and our conscious experience indeed contains gaps. This temporary change of the arrow of time would change zero energy state as a superposition of space-time surfaces analogous to almost deterministic Bohr orbits and defining a quantum goal for the system.

Maybe learning by trial and error b pairs of BSFRs leading to a temporary change of the arrow of time and demonstrating that certain BSFRs are not favourable. We might even learn moral rules and ethics ("Try to increase quantum coherence" as a basic ethical rule) by this kind of intentional trial and error process. Maybe this could also occur for computers with life-like properties.

4.5 An analog of a multi-perceptron model related to holography in TGD

TGD suggests a non-trivial example, which might have some relevance some day.

1. In TGD, holography assigns to a given 3-surface X^3 an almost unique 4-D minimal surface X^4 in 8-D space $H = M^4 \times CP_2$. X^4 can have lower-D singularities. X^4 is defined by a vanishing of 4 functions g_k of H coordinates for which TGD suggests a general form.

2. The parameters determining g_k must be determined from the a priori knowledge of the 3-surface X^3, which can be chosen to correspond to a constant value of a suitably chosen time coordinate t of M^4. X^3 takes the role of a feature which the system must learn to detect by varying the parameters appearing in the functions g_k.

3. The inputs to perceptron would be the 8 coordinates of H with t fixed. The arguments y_k of the outputs are the 4 functions g_k, which must vanish at X^4. The response functions $f_k(y_k = g_k)$ must vanish at X^3. The feedback modifies the parameters appearing in the functions g_k and the system should find the parameter values producing X^3.

4. Holography means that, apart from the failure of complete determinism, the system learns also to predict the behavior of the system as "Bohr orbit" $X^4(X^3)$ of X^3.

Acknowledgements: I want to thank our Zoom group, which includes Marko Manninen, Tuomas Sorakivi and Rode Majakka, for inspiring discussions. Special thanks go to Marko Manninen for his experimentation with chatGPT.

REFERENCES

Cosmology and Astro-Physics

[E1] Nottale L Da Rocha D. Gravitational Structure Formation in Scale Relativity, 2003. Available at: https://arxiv.org/abs/astro-ph/0310036.

Neuroscience and Consciousness

[J1] Peoch R. Chicken-robot interaction. *Medical Network.*, 62, 1995. Available at: https://paranormal.se/psi/pk/djur.html.

Books related to TGD

[K1] Pitkänen M. Construction of elementary particle vacuum functionals. In *p-Adic Physics*. Available at: https://tgdtheory.fi/pdfpool/elvafu.pdf, 2006.

[K2] Pitkänen M. Massless states and particle massivation. In *p-Adic Physics*. Available at: https://tgdtheory.fi/pdfpool/mless.pdf, 2006.

[K3] 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.

[K4] Pitkänen M. *TGD as a Generalized Number Theory: Part I.* Online book. Available at: https://www.tgdtheory.fi/tgdhtml/tgdnumber1.html., 2006.

[K5] Pitkänen M. *TGD as a Generalized Number Theory: Part II.* Online book. Available at: https://www.tgdtheory.fi/tgdhtml/tgdnumber2.html., 2006.

[K6] Pitkänen M. *p-Adic length Scale Hypothesis.* Online book. Available at: https://www.tgdtheory.fi/tgdhtml/padphys.html., 2013.

[K7] Pitkänen M. TGD as a Generalized Number Theory: p-Adicization Program. In *TGD as a Generalized Number Theory: Part I.* Available at: https://tgdtheory.fi/pdfpool/visiona.pdf, 2019.

Articles about TGD

[L1] Pitkänen M. Artificial Intelligence, Natural Intelligence, and TGD. Available at: https://tgdtheory.fi/public_html/articles/AITGD.pdf., 2017.

[L2] Pitkänen M. DMT, pineal gland, and the new view about sensory perception. Available at: https://tgdtheory.fi/public_html/articles/dmtpineal.pdf., 2017.

[L3] Pitkänen M. Philosophy of Adelic Physics. Available at: https://tgdtheory.fi/public_html/articles/adelephysics.pdf., 2017.

[L4] Pitkänen M. Philosophy of Adelic Physics. In *Trends and Mathematical Methods in Interdisciplinary Mathematical Sciences*, pages 241–319. Springer.Available at: https://link.springer.com/chapter/10.1007/978-3-319-55612-3_11, 2017.

[L5] Pitkänen M. New insights about quantum criticality for twistor lift inspired by analogy with ordinary criticality. Available at: https://tgdtheory.fi/public_html/articles/zeocriticality.pdf., 2018.

[L6] Pitkänen M. Some comments related to Zero Energy Ontology (ZEO). Available at: https://tgdtheory.fi/public_html/articles/zeoquestions.pdf., 2019.

[L7] Pitkänen M. A critical re-examination of $M^8 - H$ duality hypothesis: part I. Available at: https://tgdtheory.fi/public_html/articles/M8H1.pdf., 2020.

[L8] Pitkänen M. A critical re-examination of $M^8 - H$ duality hypothesis: part II. Available at: https://tgdtheory.fi/public_html/articles/M8H2.pdf., 2020.

[L9] Pitkänen M. Spin Glasses, Complexity, and TGD. https://tgdtheory.fi/public_html/articles/sg.pdf., 2021.

[L10] Pitkänen M. About the number theoretic aspects of zero energy ontology. https://tgdtheory.fi/public_html/articles/ZEOnumber.pdf., 2022.

[L11] Pitkänen M. Comparison of Orch-OR hypothesis with the TGD point of view. https://tgdtheory.fi/public_html/articles/penrose.pdf., 2022.

[L12] Pitkänen M. Finite Fields and TGD. https://tgdtheory.fi/public_html/articles/finitefieldsTGD.pdf., 2022.

[L13] Pitkänen M. How animals without brain can behave as if they had brain. https://tgdtheory.fi/public_html/articles/precns.pdf., 2022.

[L14] Pitkänen M. The possible role of spin glass phase and p-adic thermodynamics in topological quantum computation: the TGD view. https://tgdtheory.fi/public_html/articles/QCCC.pdf., 2022.

[L15] Pitkänen M. Two objections against p-adic thermodynamics and their resolution. https://tgdtheory.fi/public_html/articles/padmass2022.pdf., 2022.

[L16] Pitkänen M. About tessellations in hyperbolic 3-space and their relation to the genetic code . https://tgdtheory.fi/public_html/articles/tessellationH3.pdf., 2023.

[L17] Pitkänen M. Could neuronal system and even GPT give rise to a computer with a variable arrow of time? https://tgdtheory.fi/public_html/articles/GPT.pdf., 2023.

[L18] Pitkänen M. Neil Gersching's vision of self-replicating robots from TGD point of view. https://tgdtheory.fi/public_html/articles/Gersching.pdf., 2023.

[L19] Pitkänen M. New findings related to the number theoretical view of TGD. https://tgdtheory.fi/public_html/articles/M8Hagain.pdf., 2023.

[L20] Pitkänen M and Rastmanesh R. Homeostasis as self-organized quantum criticality. Available at: https://tgdtheory.fi/public_html/articles/SP.pdf., 2020.