Artifitial intelligence from first principles.

Michail Zak

Jet Propulsion Laboratory California Institute of Technology, Pasadena, CA 91109, USA

I can calculate the motion of heavenly bodies, but not the madness of people.

Isaac Newton

Abstract.

The objective of this paper is to relate the concept of intelligence to the first principles of physics, and, in particular, to answer the following question: can AI system composed only of physical components compete with a human? The answer is proven to be negative if the AI system is based only on simulations, and positive if digital devices are included. It has been demonstrated that in mathematical world, the bridge from matter to intelligence requires extension and modification of quantum physics. Such an extension is implemented by quantum neural nets that perform simulations combined with digital punctuations. The universality of this quantum-classical hybrid is in capability to violate the second law of thermodynamics by moving from disorder to order without external resources. This advanced capability that could be accepted as a definition of human-like intelligence is illustrated by examples.

1.Introduction.

The recent statement about completeness of the physical picture of our Universe made in Geneva raised many questions, and one of them is the ability to create Life and Intelligence out of physical matter without any additional entities. The main difference between living and non-living matter is in directions of their evolution: it has been recently recognized that the evolution of livings is progressive in a sense that it is directed to the highest levels of complexity. Such a property is not consistent with the behavior of *isolated* Newtonian systems that cannot increase their complexity without external forces. That difference created so called Schrödinger paradox: in a world governed by the second law of thermodynamics, all isolated systems are expected to approach a state of maximum *disorder;* since life approaches and maintains a highly *ordered* state – one can argue that this violates the Second Law implicating a paradox,[1]

But livings are not isolated due to such processes as metabolism and reproduction: the increase of order inside an organism is compensated by an increase in disorder outside this organism, and that removes the paradox. Nevertheless it is still tempting to find a mechanism that drives livings from disorder to order. The purpose of this paper is to demonstrate that moving from a disorder to order is not a prerogative of open systems: an isolated system can do it without help from outside. However such system cannot belong to the world of the modern physics: it belongs to the world of living matter, and that lead us to the concept of an intelligent particle – the first step to physics of livings. In order to introduce such a particle, we start with an idealized mathematical model of livings by addressing only one aspect of Life: a ***biosignature,*** i.e. ***mechanical*** invariants of Life, and in particular, the ***geometry and kinematics of intelligent behavior*** disregarding other aspects of Life such as metabolism and reproduction. By narrowing the problem in this way, we are able to extend the mathematical formalism of physics' First Principles to include description of intelligent behavior. At the same time, by ignoring metabolism and reproduction, we can make the system isolated, and it will be a challenge to show that it still can move from disorder to order.

2. Starting with quantum mechanics.

The starting point of our approach is the Madelung equation that is a hydrodynamics version of the Schrödinger equation

$$\frac{\partial\rho}{\partial t}+\nabla\bullet(\frac{\rho}{m}\nabla S)=0 \qquad (1)$$

$$\frac{\partial S}{\partial t}+(\nabla S)^2+F-\frac{\hbar^2\nabla^2\sqrt{\rho}}{2m\sqrt{\rho}}=0 \qquad (2)$$

Here ρ and S are the components of the wave function $\psi=\sqrt{\rho}e^{iS/\hbar}$, and $\hbar$ is the Planck constant divided by 2π. The last term in Eq. (2) is known as quantum potential. From the viewpoint of Newtonian mechanics, Eq. (1) expresses continuity of the flow of probability density, and Eq. (2) is the Hamilton-Jacobi equation for the action S of the particle. Actually the quantum potential in Eq. (2), as a feedback from Eq. (1) to Eq. (2), represents the difference between the Newtonian and quantum mechanics, and therefore, it is solely responsible for fundamental quantum properties.

The Madelung equations (1), and (2) can be converted to the Schrödinger equation using the ansatz

$$\sqrt{\rho}=\Psi\exp(-iS/h) \qquad (3)$$

where ρ and S being real function.

Our approach is based upon a modification of the Madelung equation, and in particular, upon replacing the quantum potential with a different Liouville feedback, Fig.1

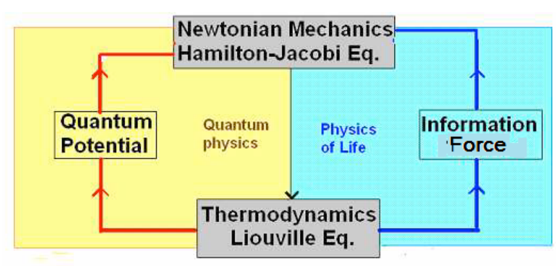

Figure 1. Classic Physics, Quantum Physics and Physics of Life.

In Newtonian physics, the concept of probability ρ is introduced via the Liouville equation

$$\frac{\partial\rho}{\partial t}+\nabla\bullet(\rho\mathbf{F})=0 \qquad (4)$$

generated by the system of ODE

$$\frac{d\mathbf{v}}{dt} = \mathbf{F}[\mathbf{v}_1(t),...\mathbf{v}_n(t),t] \tag{5}$$

where **v** is velocity vector.
It describes the continuity of the probability density flow originated by the error distribution

$$\rho_0 = \rho(t=0) \tag{6}$$

in the initial condition of ODE (6).

Let us rewrite Eq. (2) in the following form

$$\frac{d\mathbf{v}}{dt} = \mathbf{F}[\rho(\mathbf{v})] \tag{7}$$

where **v** is a velocity of a hypothetical particle.

This is a fundamental step in our approach: in Newtonian dynamics, the probability never explicitly enters the equation of motion, [2,3]. In addition to that, the Liouville equation generated by Eq. (7) is nonlinear with respect to the probability density ρ

$$\frac{\partial \rho}{\partial t} + \nabla \bullet \{\rho \mathbf{F}[\rho(\mathbf{V})]\} = 0 \tag{8}$$

and therefore, the system (7),(8) departs from Newtonian dynamics. However although it has the same topology as quantum mechanics (since now the equation of motion is coupled with the equation of continuity of probability density), it does not belong to it either. Indeed Eq. (7) is more general than the Hamilton-Jacoby equation (2): it is not necessarily conservative, and **F** is not necessarily the quantum potential although further we will impose some restriction upon it that links **F** to the concept of information, [3]. The relation of the system (7), (8) to Newtonian and quantum physics is illustrated in Fig.1.

Remark. Here and below we make distinction between the random *variable v(t)* and its *values V* in probability space.

3. Information force instead of quantum potential.

In this section we propose the structure of the force **F** that plays the role of a feedback from the Liouville equation (8) to the equation of motion (7). Turning to one-dimensional case, let us specify this feedback as

$$F = c_0 + \frac{1}{2}c_1\rho - \frac{c_2}{\rho}\frac{\partial \rho}{\partial v} + \frac{c_3}{\rho}\frac{\partial^2 \rho}{\partial v^2} \tag{9}$$

$$c_0 > 0, c_1 > 0, c_3 > 0 \tag{10}$$

Then Eq.(9) can be reduced to the following:

$$\dot{v} = c_0 + \frac{1}{2}c_1\rho - \frac{c_2}{\rho}\frac{\partial \rho}{\partial v} + \frac{c_3}{\rho}\frac{\partial^2 \rho}{\partial v^2} \tag{11}$$

and the corresponding Liouville equation will turn into the nonlinear PDE

$$\frac{\partial \rho}{\partial t} + (c_0 + c_1\rho)\frac{\partial \rho}{\partial V} - c_2\frac{\partial^2 \rho}{\partial v^2} + c_3\frac{\partial^3 \rho}{\partial V^3} = 0 \tag{12}$$

This equation is known as the KdV-Bergers' PDE. The mathematical theory behind the KdV equation became rich and interesting, and, in the broad sense, it is a topic of active mathematical research. A homogeneous version of this equation that illustrates its distinguished properties is nonlinear PDE of parabolic type. But a fundamental difference between the standard KdV-Bergers equation and Eq. (12) is that Eq. (12) ***dwells in the probability space***, and therefore, it must satisfy the normalization constraint

$$\int_{-\infty}^{\infty} \rho \, dV = 1 \tag{13}$$

However as shown in [4], this constraint is satisfied: in physical space it expresses conservation of mass, and it can be easily scale-down to the constraint (13) in probability space. That allows one to apply all the known results directly to Eq. (12). However it should be noticed that all the conservation invariants have different physical meaning: they are not related to conservation of momentum and energy, but rather impose constraints upon the Shannon information.

In physical space, Eq. (12) has many applications from shallow waves to shock waves and solitons. However, application of solutions of the same equations in probability space is fundamentally different. In the following sections we will present a phenomena that exist neither in Newtonian nor in quantum physics.

4. Emergence of randomness.

In this section we discuss a fundamentally new phenomenon: transition from determinism to randomness in ODE that coupled with their Liouville PDE.

In order to complete the solution of the system (11), (12), one has to substitute the solution of Eq. (12):

$$\rho = \rho(V, t) \quad at \quad V = v \tag{14}$$

into Eq.(11). Since the transition from determinism to randomness occurs at $t \to 0$, let us turn to Eq. (12) with sharp initial condition

$$\rho_0(V) = \delta(V) \, at \qquad t = 0, \tag{15}$$

Then applying one of the standard analytical approximations of the delta-function, one obtains the asymptotic solution

$$\rho = \frac{1}{t\sqrt{\pi}} e^{-\frac{V^2}{t^2}} \quad at \quad t \to 0 \tag{16}$$

Substitution this solution into Eq. (14) shows that

$$\begin{gathered} O(c_0 + \frac{1}{2}c_1\rho) = \frac{1}{t}, \quad O(\frac{c_2}{\rho}\frac{\partial \rho}{\partial v}) = \frac{1}{t^2}, \\ and \quad O(\frac{c_3}{\rho}\frac{\partial^2 \rho}{\partial v^2}) = \frac{1}{t^4} \quad at \quad t \to 0, \quad v \neq 0 \end{gathered} \tag{17}$$

i.e.

$$c_0 + \frac{1}{2}c_1\rho << \frac{c_2}{\rho}\frac{\partial \rho}{\partial v} << \frac{c_3}{\rho}\frac{\partial^2 \rho}{\partial v^2} \quad at \quad t \to 0, \quad v \neq 0 \tag{18}$$

and therefore, the first three terms in Eq. (11) can be ignored

$$\dot{v} = \frac{c_3}{\rho}\frac{\partial^2 \rho}{\partial v^2} \quad at \quad t \to 0, \quad v \neq 0 \tag{19}$$

or after substitution of Eq. (16)

$$\dot{v} = \frac{4c_3 v^2}{t^4} \quad at \quad t \to 0, \quad v \neq 0 \tag{20}$$

Eq. (20) has the following solution (see Fig. 2)

$$v = \frac{t^3}{4c_3 + Ct^3} \quad at \quad t \to 0, \quad v \neq 0 \tag{21}$$

where C is an arbitrary constant.

This solution has the following property: the Lipchitz condition at $t \to 0$ fails

$$\frac{\partial \dot{v}}{\partial v} = \frac{8c_3 v}{t^4} = \frac{8c_3 t^3}{t^4(4c_3 + Ct^3)} \to \infty \quad at \quad t \to 0, \quad v \neq 0 \tag{22}$$

and as a result of that, the uniqueness of the solution is lost. Indeed, as follows from Eq. (21), for any value of the arbitrary constant C, the solutions are different, but they satisfy the same initial condition

$$v \to 0 \quad at \quad t \to 0 \tag{23}$$

Due to violation of the Lipchitz condition (22), the solution becomes unstable. That kind of instability when infinitesimal errors lead to finite deviations from basic motion (the Lipchitz instability) has been discussed in [5,6]. This instability leads to unpredictable shift of solution from one value of *C* to another. It means that appearance of any specified solution out of the whole family is random, and that randomness is controlled by the feedback (9) from the Liouville equation (12). Indeed if the solution (21) runs independently many times with the same initial conditions, and the statistics is collected, the probability density will satisfy the Liouville equation (12), Fig.3.

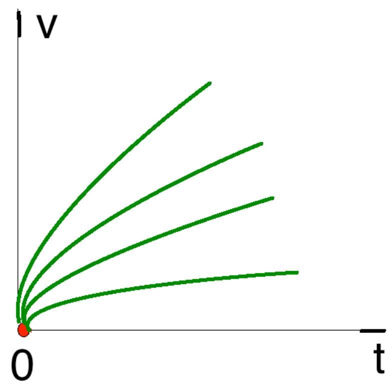

Figure 2. Family of random solutions describing transition from determinism to stochastisity.

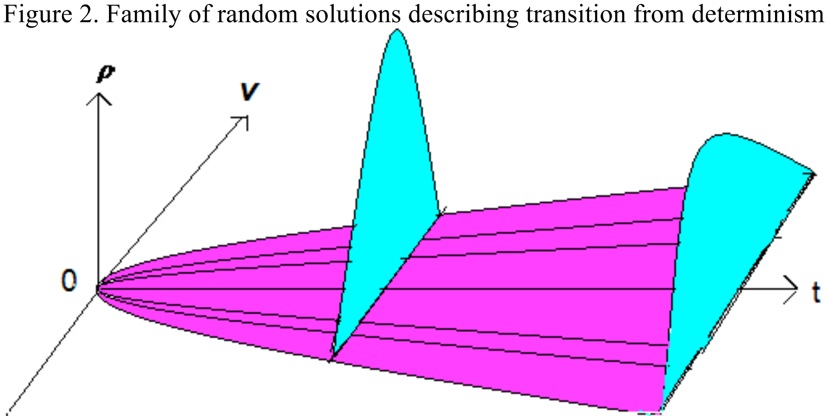

Figure 3. Stochastic process and probability density.

5. Departure from Newtonian and quantum physics.

In this section we will derive a distinguished property of the system (16),(17) that is associated with violation of the second law of thermodynamics i.e. with the capability of moving from disorder to order without help from outside. That property can be predicted qualitatively even prior to analytical proof: due to the nonlinear term in Eq. (17), the solution form shock waves and solitons in probability space, and that can be interpreted as "concentrations" of probability density, i.e. departure from disorder. In order to demonstrate it analytically, let us turn to Eq. (17) at

$$c_1 >> |c_2|, c_3 \tag{24}$$

and find the change of entropy H

$$\frac{\partial H}{\partial t} = -\frac{\partial}{\partial t}\int_{-\infty}^{\infty} \rho \ln \rho \, dV = -\int_{-\infty}^{\infty} \frac{1}{c_1}\dot{\rho}(\ln \rho + 1)\, dV = \int_{-\infty}^{\infty} \frac{1}{c_1}\frac{\partial}{\partial V}(\rho^2)\ln(\rho+1)\, dV$$
$$= \frac{1}{c_1}[\, \Big|_{-\infty}^{\infty} \rho^2(\ln \rho + 1) - \int_{-\infty}^{\infty} \rho \, dV] = -\frac{1}{c_1} < 0 \tag{25}$$

At the same time, the original system (11), (12) is isolated: it has no external interactions. Indeed the information force Eq. (9) is generated by the Liouville equation that, in turn, is generated by the equation of motion (11). In addition to that, the particle described by ODE (11) is in equilibrium $\dot{v} = 0$ prior to activation of the feedback (9). Therefore the solution of Eqs. (11), and (12) can violate the second law of thermodynamics, and that means that this class of dynamical systems does not belong to physics as we know it. This conclusion triggers the following question: are there any phenomena in Nature that can be linked to dynamical systems (11), (12)? The answer will be discussed bellow.

Thus despite the mathematical similarity between Eq.(12) and the KdV-Bergers equation, the physical interpretation of Eq.(12) is fundamentally different: it is a part of the dynamical system (11),(12) in which Eq. (12) plays the role of the Liouville equation generated by Eq. (11). As follows from Eq. (25), this system, being isolated and being in equilibrium, has the capability to decrease entropy, i.e. to move from disorder to order without external resources. In addition to that, the system displays transition from deterministic state to randomness (see Eq. (22)).

This property represents departure from classical and quantum physics, and, as shown in [2,3], provides a link to behavior of livings. That suggests that this kind of dynamics requires extension of modern physics to include physics of life.

The process of violation of the second law of thermodynamics is illustrated in Fig. 4: the higher values of ρ propagate faster than lower ones. As a result, the moving front becomes steeper and steeper, and that leads to formation of solitons (c_3>0), or shock waves (c_3=0) ***in probability space.*** This process is accompanied by decrease of entropy.

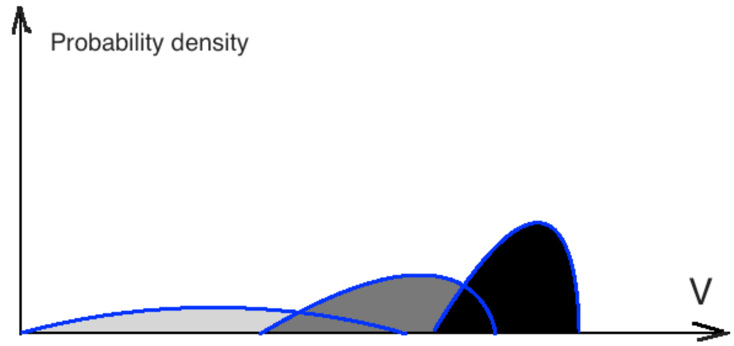

Figure 4. Formation of shock waves in probability space.

Remark. The system (11), (12) displays transition from deterministic state to randomness (see Eq. (22)), and this property can be linked to the similar property of the Madelung equation, although strictly speaking, Eq.(1) is a "truncated" version of the Liouvile equation: it does not include the contribution of the quantum potential. Nevertheless the origin of randomness in quantum mechanics is the same as in the system (11), (12) as demonstrated in [3,6].

6. Comparison with quantum mechanics.

a. Mathematical Viewpoint. The model of intelligent particle is represented by a nonlinear ODE (7) and a nonlinear parabolic PDE (8) coupled in a master-slave fashion: Eq. (8) is to be solved independently, prior to solving Eq. ((7). The coupling is implemented by a feedback that includes the probability density and its space derivatives, and that converts the first order PDE (the Liouville equation) to the second or higher order nonlinear PDE. As a result of the nonlinearity, the solutions to PDE can have attractors (static, periodic, or chaotic) in probability space. The solution of ODE (7) represents another major departure from classical ODE: due to violation of Lipchitz conditions at states where the probability density has a sharp value, the solution loses its uniqueness and becomes random. However, this randomness is controlled by the PDE (8) in such a way that each random sample occurs with the corresponding probability, Fig.3.

b. Physical Viewpoint. The model of intelligent particle represents a fundamental departure from both Newtonian and quantum mechanics. The fundamental departure of all the modern physics is the violation of the second laws of thermodynamics,(see Eq.(25), and Fig. 4). However a more detailed analysis, [3], shows that due to similar dynamics topology to quantum mechanics,(see Fig.1) the model preserves some quantum properties such as entanglement and interference of probabilities.

7. Origin of intelligence.

a. Relevance to model of intelligent particle. The proposed model illuminates the "border line" between living and non-living systems. The model introduces an intelligent particle that, in addition to Newtonian properties, possesses the ability to process information. The probability density can be associated with the *self-image* of the intelligent particle as a member of the class to which this particle belongs, while its ability to convert the density into the information force - with the *self-awareness* (both these concepts are adopted from psychology). Continuing this line of associations, the equation of motion (such as Eq (11)) can be identified with a motor dynamics, while the evolution of density (see Eq. (12)) –with a mental dynamics. Actually the mental dynamics plays the role of the Maxwell sorting demon: it rearranges the probability distribution by creating the information potential and converting it into a force that is applied to the particle. One should notice that mental dynamics describes evolution of the whole class of state variables (differed from each other only by initial conditions), and that can be associated with the ability to generalize that is a privilege of intelligent systems. Continuing our biologically inspired interpretation, it should be recalled that the second law of thermodynamics states that the entropy of an isolated system can only increase. This law has a clear probabilistic interpretation: increase of entropy corresponds to the passage of the system from less probable to more probable states, while the highest probability of the most disordered state (that is the state with the highest entropy) follows from a simple combinatorial analysis. However, this statement is correct only if there is no Maxwell' sorting demon, i.e., nobody inside the system is rearranging the probability distributions. But this is precisely what the Liouville feedback is doing: it takes the probability density ρ from Equation (12), creates functions of this density, converts them into the information force and applies this force to the equation of motion (11). As already mentioned above, because of that property of the model, the evolution of the probability density can become nonlinear, and the entropy may decrease "against the second law of thermodynamics". Actually the proposed model represents governing equations for interactions of intelligent agents. In order to emphasize the autonomy of the agents' decision-making process, we will associate the proposed models with ***self-supervised (SS) active systems.*** By an active system we will understand here a set of interacting intelligent agents capable of processing information, while an intelligent agent is an autonomous entity, which observes and acts upon an environment and directs its activity towards achieving goals. The active system is not derivable from the Lagrange or Hamilton principles, but it is rather created for information processing. One of specific differences between active and physical systems is that the former are supposed to act in uncertainties originated from incompleteness of information. Indeed, an intelligent agent almost never has access to the whole truth of its environment. Uncertainty can also arise because of incompleteness and incorrectness in the agent's understanding of the properties of the environment. That is why *quantum-inspired SS* systems represented by the particles under consideration are well suited for representation of active systems, and the hypothetical particle introduced above can be associated with the term "intelligent" particle. It is important to emphasize that self-supervision is implemented by the feedback from mental dynamics, i.e. by internal force, since the mental dynamics is generated by intelligent particle itself.

b. Comparison with control systems. In this sub-section we will establish a link between the concepts of intelligent control and phenomenology of behavior of intelligent particle.

Example. One of the limitations of classical dynamics, and in particular, neural networks, is inability to change their structure without an external input. As will be shown below, an intelligent particle can change the locations and even the type of the attractors being triggered only by information forces i.e. by an internal effort. We will start with a simple dynamical system

$$\dot{v} = 0, \quad v = 0 \quad at\ t = 0 \tag{26}$$

and than apply the following control

$$F = -k\overline{v} + a\overline{\overline{v}} - \sigma \frac{\partial}{\partial v} \ln \rho, \tag{27}$$

where $$\overline{\overline{V}} = \int_{-\infty}^{\infty} \rho (V - \overline{V})^2 dV, \qquad \overline{V} = \int_{-\infty}^{\infty} \rho V dV, \tag{28}$$

and k, a, σ are constant coefficients.

Then the controlled version of the motor dynamics (26) is changed to

$$\dot{v} = -k\overline{v} + a\overline{\overline{v}} - \sigma \frac{\partial}{\partial v} \ln \rho \tag{29}$$

while F represents the information forces that play the role of ***internal*** actuator.

Let us notice that the internal actuator (27) is a particular case of the information force (9) at

$$c_0 = -k\overline{v} + a\overline{\overline{v}}, \quad c_1 = 0, \quad c_2 = \sigma, \quad c_3 = 0 \tag{30}$$

For a closure, Eq. (29) is complemented by the corresponding Liouville equation

$$\frac{\partial\rho}{\partial t} = k\overline{V}\frac{\partial\rho}{\partial V} - a\overline{\overline{V}}\frac{\partial\rho}{\partial V} + \sigma\frac{\partial^2\rho}{\partial V^2}, \tag{31}$$

to be solved subject to sharp initial condition

$$\rho_0(V) = \delta(V)\, at \qquad t = 0, \tag{32}$$

As shown above, the solution of Eq.(29) is random, (see Eq. (21) and Fig. 2) while this randomness is controlled by Eq. (31). Therefore in order to describe it, we have to transfer to the mean values $\overline{v}$ and $\overline{\overline{v}}$. For that purpose, let us multiply Eq.(1) by V .Then integrating it with respect to V over the whole space, one arrives at ODE for the expectation $\overline{v}(t)$

$$\dot{\overline{v}} = -k\overline{v} + a\overline{\overline{v}} \tag{33}$$

Multiplying Eq.(31) by V^2, then integrating it with respect to V over the whole space, one arrives at ODE for the variance $\overline{\overline{v}}(t)$

$$\dot{\overline{\overline{v}}} = -2k\overline{\overline{v}} + 2a\overline{v}\,\overline{\overline{v}} + 2\sigma \tag{34}$$

Let us find fixed points of the system (33) and (34) by solving the system of algebraic equations:

$$0 = -k\overline{v} + a\overline{\overline{v}} \tag{35}$$

$$0 = -2k\overline{\overline{v}} + 2a\overline{v}\,\overline{\overline{v}} + 2\sigma \tag{36}$$

By selecting

$$\sigma = \frac{k^3}{2a^2} \tag{37}$$

we arrive at the following single fixed point

$$\overline{v}^* = \frac{k}{2a}, \quad \overline{\overline{v}}^* = \frac{k^2}{2a^2} \tag{38}$$

In order to establish whether this fixed point is an attractor or a repeller, we have to analyze stability of the homogeneous version of the system (33), (34) linearized with respect to the fixed point (38)

$$\dot{\overline{v}} = -k\overline{v} + a\overline{\overline{v}} \tag{39}$$

$$\dot{\overline{\overline{v}}} = -k\overline{\overline{v}} + \frac{k^2}{a}\overline{v} \tag{40}$$

Analysis of its characteristic equation shows that it has non-positive roots:

$$\lambda_1 = 0, \quad \lambda_2 = -2k < 0$$

and therefore, the fixed point (38) is a stochastic attractor with stationary mean and variance. However the higher moments of the probability density are not necessarily stationary: they can be found from the original PDE (31).

Thus as a result of a ***mental*** control, an ***isolated*** dynamical system (26) that prior to control was at rest, moves to the stochastic attractor (38) having the expectation $\overline{v}*$ and the variance $\overline{\overline{v}}*$.

The distinguished property of the particle introduced above definitely fits into the concept of intelligence. Indeed, the evolution of intelligent living systems is directed toward the highest levels of complexity if the complexity is measured by an irreducible number of different parts that interact in a well-regulated fashion. At the same time, the solutions to the models based upon dissipative Newtonian dynamics eventually approach attractors where the evolution stops while these attractors dwell on the subspaces of lower dimensionality, and therefore, of the lower complexity (until a "master" reprograms the model). Therefore, such models fail to provide an autonomous progressive evolution of intelligent systems (i.e. evolution leading to increase of complexity). At the same time, a self-controlled particle can create its own complexity based only upon an *internal* effort.

Thus the actual source of intelligent behavior of the particle introduced above is a new type of force - the information force - that contributes its work into the Law of conservation of energy. However this force is internal: it is generated by the particle itself with help of the Liouvile equation. The machinery of the intelligence is similar to that of control system with the only difference that control systems are driven by external actuators while the intelligent particle is driven by a feedback from the Liouvile equation without any external resources.

8. Quantum recurrent neural nets.

a. Background.

In the previous sections, we presented a ***mathematical*** answer to the ancient philosophical question "How mind is related to matter". We proved that in mathematical world, the bridge from matter to mind requires extension and modification of quantum physics. In this context, we will comment on the recent statement made by Stephen Hawking on December 2, 2014, in which he warns that artificial intelligence could end mankind. Based upon our work, part of which is presented in the previous sections, it can be stated that machines composed only out of physical components and ***without any digital devices*** being included, cannot, in principle, overperform a human in creativity, regardless of the level of technology. But what happens if a machine does include digital devices? The answer to this question is the subject of the following sections. In these sections we propose a quantum version of recurrent neural nets (QRN) that along with classical performance, possess the capability to move from disorder to order without external recourses, and that makes their intelligence comparable with that of a human. The QRN incorporate classical feedback loops into conventional quantum networks. It is shown, [7],[11], that dynamical evolution of such networks, which interleave quantum evolution with measurement and reset operations, exhibits novel dynamical properties. Moreover, decoherence in quantum recurrent networks is less problematic than in conventional quantum network architectures due to the modest phase coherence times needed for network operation. It is proven that a hypothetical quantum computer can implement an exponentially larger number of the degrees of freedom within the same size.

b. Quantum model of evolution.

A state of a quantum system is described by a special kind of time dependent vector $|\psi>$ with complex components called amplitudes:

$$\{a_0 a_1 ... a_n\} = |\psi> \tag{41}$$

If unobserved, the state evolution is governed by the Schrödinger equation:

$$i\hbar \frac{da_k}{dt} = \sum_l H_{kl} a_l \tag{42}$$

which is linear and reversible.

Here H_{kl} is the Hamiltonian of the system, $i = \sqrt{-1}$, $\hbar = 1.0545 \times 10^{-34} JS$.

The solution of Eq. (42) can be written in the following form:

$$\{a_0(t),..a_n(t)\} = \{a_0(0),..a_n(0)\}U* \tag{43}$$

where U is a unitary matrix uniquely defined by the Hamiltonian:

$$U = e^{-iHt/\hbar}, \qquad UU* = I \tag{44}$$

After m equal time steps Δt

$$\{a_0(m\Delta t),...a_n(m\Delta t)\} = \{a_0(0),...a_n(0)\}U^{*m} \qquad (45)$$

the transformation of the amplitudes formally looks like those of the transition probabilities in Markov chains. However, there is a fundamental difference between these two processes: in Eq. (45) the probabilities are represented not by the amplitudes, but by squares of their modules:

$$p = \{|a_0|^2,...|a_n|^2\} \qquad (46)$$

and therefore, the unitary matrix U is not a transition probability matrix.
It turns out that this difference is the source of so called quantum interference, which makes quantum computing so attractive. Indeed, due to interference of quantum probabilities:

$$p = |a_1 + a_2|^2 \neq p_1 + p_2 \qquad (47)$$

each element of a new vector $a_i(m\Delta t)$ in Eq. (45) will appear with the probability $|a_i|^2$ that includes all the combinations of the amplitudes of the previous vector.

c. Quantum Collapse and Sigmoid Function.

As well known, neural nets have two universal features: dissipativity and nonlinearity. Due to dissipativity, a neural net can converge to an attractor and this convergence is accompanied by a loss of information. But such a loss is healthy: because of it, a neural net filters out insignificant features of a pattern vector while preserving only the invariants which characterizes it's belonging to a certain class of patterns. These invariants are stored in the attractor, and therefore, the process of convergence performs generalization: two different patterns that have the same invariants will converge to the same attractor. Obviously, this convergence is irreversible. The nonlinearity increases the neural net capacity: it provides many different attractors including static, periodic, chaotic and erogdic, and that allows one to store simultaneously many different patterns. Both dissipativity and nonlinearity are implemented in neural nets by the sigmoid (or squashing) function. It is important to emphasize that the only qualitative properties of the sigmoid function are those, which are important for the neural net performance, but not any specific forms of this function**.** Can we find a qualitative analog of a sigmoid function in quantum mechanics? Fortunately, yes: it is so called quantum collapse that occurs as a result of quantum measurements. Indeed, the result of any quantum measurement is always one of the eigenvalues of the operator corresponding to the observable being measured. In other words, a measurement maps a state vector of the amplitudes (41) into an eigenstate vector

$$\{a_0 a_1 ... a_n\} \rightarrow \{00...\underset{\uparrow_i}{1}...00\} \qquad (48)$$

while the probability that this will be the i^{th} eigenvector is:

$$p_i = |a_i|^2 \qquad (49)$$

The operation (49) is nonlinear, dissipative, and irreversible, and it can play the role of a natural "quantum" sigmoid function.

10. QRN Architectures.

Let us introduce the following sequence of transformations for the state vector (41):

$$|\psi(0)> \rightarrow U|\psi(0)> \rightarrow \sigma_1\{U|\psi(0)|\} = |\psi(t+1)> \qquad (50)$$

which is a formal representation of Eq.(48)) with σ_1 denoting a "quantum" sigmoid function.

In order to continue this sequence, we have to reset the quantum device considering the resulting eigenstate as a new input. Then one arrives at the following neural net:

$$a_i(t+1) = \sigma_1\{\sum U_{ij} a_j(t)\}, \qquad i = 1.2...n \qquad ((51)$$

The curly brackets are intended to emphasize that σ_1 is to be taken as a measurement operation with the effect similar to those of a sigmoid function in classical neural networks (Fig. 5).

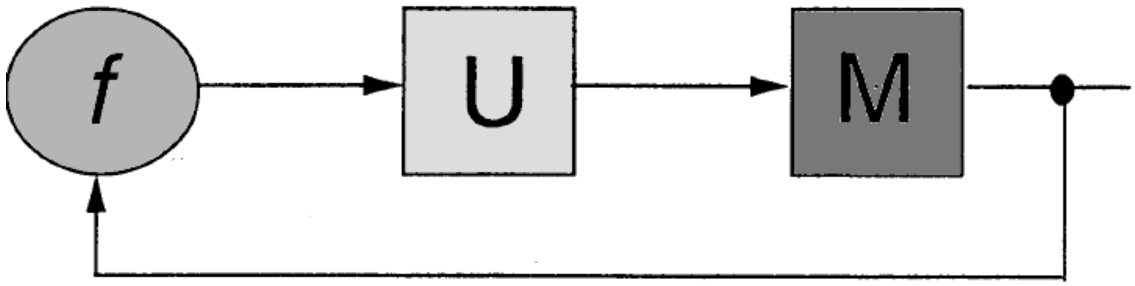

Figure 5. The simplest architecture of quantum neural net.

However, there are two significant differences between the quantum (51) and classical neural nets. Firstly, in Eq. (51) the randomness appears in the form of quantum measurements as a result of the probabilistic nature of the quantum mechanics, while in neural network a special device generating random numbers is required. Secondly, if the dimension of the classical matrix T_{ij} is $N \times N$, then within the same space one can arrange the unitary matrix *U* (or the Hamiltonian *H*) of dimension $2^N \times 2^N$ exploiting the quantum entanglement and direct product decomposability of the Schrödinger equation. One should notice that each non-diagonal element of the matrix *H* might consist of two independent components: real and imaginary. The only constraint imposed upon these elements is that *H* is the Hermitian matrix, i.e.,

$$H_{ij} = \overline{H}_{ji} \tag{52}$$

and therefore, the $n \times n$ Hermitian matrix has n^2 independent components.

So far the architecture of the neural net (51) was based upon one measurement per each run of the quantum device. However, in general, one can repeat each run for *l* times $l \leq n$ collecting *l* independent measurements. Then, instead of the mapping (48), one arrives at the following best estimate of the new state vector:

$$\{a_0 ... a_n\} \rightarrow \{0 ... \underset{\uparrow_{i_1}}{\frac{1}{\sqrt{l}}} ... 0 ... \underset{\uparrow_{i_l}}{\frac{1}{\sqrt{l}}} ...\} \tag{53}$$

while the probability that the new state vector has non-zero i_k^{th} component is

$$p_{ik} = |a_{ik}|^2 \tag{54}$$

Denoting the sigmoid function corresponding to the mapping (53) as σ_l, one can rewrite Eq. (51) in the following form:

$$a_i(t+1) = \sigma_l\{\sum U_{ij} a_j(t)\}, \qquad i = 1.2...n \tag{55}$$

The next step in complexity of the ORN architecture can be obtained if one introduces several quantum devices with synchronized measurements and resets:

$$a_i^{(1)}(t+1) = \sigma_{l_1 l_2}\{\sum U_{ij}^{(1)} a_j^{(1)}(t)\}, \quad i = 1.2...n_1 \tag{56}$$

$$a_i^{(2)}(t+1) = \sigma_{l_2 l_1}\{\sum U_{ij}^{(2)} a_j^{(2)}(t)\}, \qquad i \neq 1.2...n_2 \tag{57}$$

Here the sigmoid functions $\sigma_{l_1 l_2}$ and $\sigma_{l_2 l_1}$, map the state vectors into a weighted mixtures of the measurements:

$$\{a_1^{(1)} ... a_n^{(1)}\} \rightarrow \frac{a_{11} a_{l_1}^{(1)} + a_{12} a_{l_2}^{(2)}}{|a_{11} a_{l_1}^{(1)} + a_{12} a_{l_2}^{(2)}|} \tag{58}$$

$$\{a_1^{(2)}...a_n^{(2)}\} \to \frac{a_{21}a_{l_1}^{(1)} + a_{22}a_{l_2}^{(2)}}{|a_{21}a_{l_1}^{(1)} + a_{22}a_{l_2}^{(2)}|} \tag{59}$$

where $a_{l_1}^{(1)}$ and $a_{l_2}^{(2)}$ are the result of measurements presented in the form (53), and a_{11}, a_{12}, a_{21} and a_{22} are constants.

Thus, Eqs. (56) and (57) evolve independently during the quantum regime, i.e., in between two consequtive measurements; however, during the measurements and resets they are coupled via the Eqs. (58) and (59). It is easy to calculate that the neural nets (51), (55) and (56), (57) operate with patterns whose dimensions are $n, n(n-1)(n-l), n_1(n_1-1)(n_1-l), n_2(n_2-1)(n_2-l)$, respectively.

In a more general architecture, one can have K-parallel quantum devices U_i with l_i consequtive measurements M_i for each of them *(i=1,2...k)*, see Fig. 2.

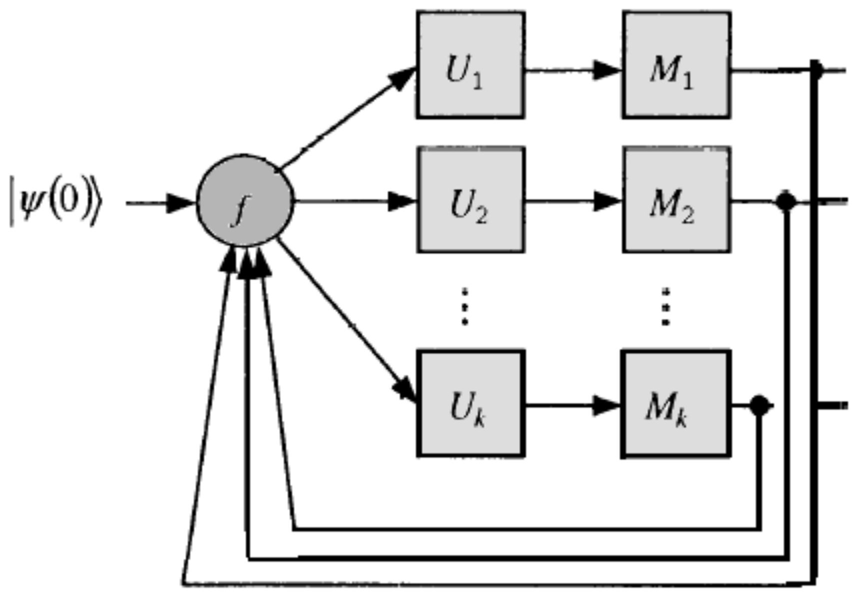

Figure 6. ***The k-Parallel Quantum Neural Network Architecture***

Recall that one is free to record, duplicate or even monitor the sequence of measurement outcomes, as they are all merely bits and hence constitute classical information. Moreover, one is free to choose the function used during the reset phase, including the possibility of adding no offset state whatsoever. Such flexibility makes the QRN architecture remarkably versatile. To simulate a Markov process, it is sufficient to return just the last output state to the next input at each iteration.

11. Evolution of probabilities.

Let us take another look at Eq. (51). Actually it performs a mapping of an i^{th} eigenvector into an j^{th} eigenvector:

$$\{00...010...0\} \to \{00...010...0\} \tag{60}$$

$$\uparrow_i \qquad\qquad \uparrow_j$$

The probability of the transition (60) is uniquely defined by the unitary matrix *U*:

$$p_{ij} = |U_{ji}|^2, \qquad \sum_{i=1}^{n} p_{ij} = 1 \tag{61}$$

and therefore the matrix $\| p_{ij} \|$ plays the role of the transition matrix in a generalized random walk which is represented by the chain of mapping (60).

Thus, the probabilistic performance of Eq. (51) has remarkable features: it is quantum (in a sense of the interference of probabilities) in between two consecutive measurements, and it is classical in description of the sequence of mapping (1). Applying the transition probability matrix (61) and starting, for example, with eigenstate $\{10...0\}$, one obtains the following sequence of the probability vectors:

$$\pi_0 = \{10...0\}; \quad \pi_1 = \{10...0\}\begin{pmatrix} p_{11}\cdots p_{11} \\ .... \\ p_{n1}\cdots p_{nn} \end{pmatrix} = \{\pi_1{}^1 ... \pi_n{}^1\}; etc \qquad (62)$$

An i^{th} component of the vector $\pi_m, i.e \quad \pi_m{}^i$ expresses the probability that the system is in the i^{th} eigenstate after m steps. As follows from Eqs. (62), the evolution of probabilities is a linear stochastic process, although each particular realization of the solution to Eq. (51) evolves nonlinearly, and one of such realization is the maximum likelihood solution. In this context, the probability distribution over different particular realizations can be taken as a measure of possible deviations from the best estimate solution. However, the stochastic process (62) as an ensemble of particular realizations, has its own invariant characteristics which can be expressed independently on these realizations. One of such characteristics is the probability $f_{ij}{}^{(m)}$ that the transition from the eigenstate *i* to the eigenstate *j* is performed in *m* steps. This characteristic is expressed via the following recursive relationships, [8]:

$$\begin{aligned} &f_{ij}{}^{(1)} = p_{ij}{}^{(1)} = p_{ij}, f_{ij}{}^{(2)} = p_{ij}{}^{(2)} - f_{ij}{}^{(1)} p_{ij} \\ &f_{ij}{}^{(m)} = p_{ij}{}^{(m)} - f_{ij}{}^{(1)} p_{jj}{}^{(n-1)} - f_{ij}{}^{(3)} p_{jj}{}^{(n-2)} ... - f_{ij}{}^{(n-1)} p_{jj}. \end{aligned} \qquad (63)$$

If

$$\sum_{m=1}^{\infty} f_{ij}{}^{(m)} < 1 \qquad (64)$$

then the process initially in the eigenstate *i* may never reach the eigenstate *j*.

If $$\sum_{m=1}^{\infty} f_{ij}{}^{(m)} = 1 \qquad (65)$$

then the i^{th} eigenstate is a recurrent state, i.e., it can be visited more than once. In partiicular, if

$$p_{ii} = 1 \qquad (66)$$

this recurrent state is an absorbing one: the process will never leave it once it enters.

From the viewpoint of neural net performance, any absorbing state represents a deterministic static attractor without a possibility of "leaks." In this context, a recurrent, but not absorbing state can be associated with a periodic or an aperiodic (chaotic) attractor. To be more precise, an eigenstate *I* has a period τ $(\tau > 1)$ *if* $p_{ii}{}^{(m)} = 0$ whenever *m* is not divisible by τ, and τ is the largest integer with this property. The eigenstate is aperiodic

if $\tau = 1$ (67)

Another invariant characteristic which can be exploited for categorization and generalization is reducibility, i.e., partitioning of the states of a Markov chain into several disjoint classes in which motion is trapped. Indeed, each hierarchy of such classes can be used as a set of filters, which are passed by a pattern before it arrives at the smallest irreducible class whose all states are recurrent. For the purpose of evaluation of deviations (or "leaks") from the maximum likelihood solution, long-run properties of the evolution of probabilities (62) are important. Some of these properties are known from theory of Markov chains, namely: for any irreducible erogodic Markov chain the limit $p_{ij}^{(m)}$ exists and it is independent of *I*, .e.,

$$\lim p_{ij}^{(m)} = \pi_i \quad at \quad m \to \infty \tag{68}$$

while

$$\pi_j > 0, \quad \pi_j = \sum_{i=0}^{k} \pi_i p_{ij}, \quad j = 0,1,..k, \quad \sum_{j=}^{k} \pi_j = 1, \quad \pi_j = \frac{1}{\mu_{ii}} \tag{69}$$

Here μ_{ii} is the expected recurrence time

$$\mu_{ii} = 1 + \sum_{l \neq j} p_{ij} \mu_{li} < \infty \tag{70}$$

The definition of ergodicity of a Markov chain is based upon the conditions for aperiodicity (67) and positive recurrence (68), while the condition for irreducibility requires existence of a value of *m* not dependent upon *i* and *j* for which $p_{ij}^{(m)}$ >0 for *i* and *j*. The convergence of the evolution (62) to a stationary stochastic process suggests additional tools for information processing. Indeed, such a process for *n*-dimensional eigenstates can be uniquely defined by *n* statistical invariants (for instance, by first *n* moments) which are calculated by summations over time rather than over the ensemble, and for that a single run of the quantum net (51) is sufficient. Hence, triggered by a simple eigenstate, a prescribed by *n*-invariants stochastic process can be retrieved and displayed for the purposes of Monte-Carlo computations, for modelling and prediction of behavior of stochastic systems, etc.

Continuing analysis of evolution of probability, let us introduce the following difference equation

$$\pi_i(t+\tau) = \sum_{j=1}^{n} \pi_j(t) p_{ij}, \quad \sum_{i=1}^{n} \pi_i = 1, \quad \pi_i \geq 0, \quad i = 1,2,..n \tag{71}$$

It should be noticed that the vector $\pi = (\pi_1,..\pi_n)$ as well as the stochastic matrix p_{ij} exist only in an abstract Euclidean space: they never appear explicitly in physical space. The evolution (71) is also irreversible, but it is linear and deterministic.

The only way to reconstruct the probability vector $\pi(t)$ is to utilize the measurement results for the vector *a(t).* In general case, many different realizations of Eq. (60) are required for that purpose. However, if the condition (64) holds, the ergodic attractor $\pi = \pi^{\infty}$ can be found from the only one realization due to the ergodicity of the stochastic process. The ergodic attractor π^{∞} can be found analytically from the steady-state equations:

$$\pi_i^{\infty} = \sum_{j=1}^{n} p_{ij} \pi_j^{\infty}, \quad \sum_{i=1}^{n} \pi_i^{\infty} = 1, \quad \sum_{j=1}^{n} p_{ij} = 1, \quad \pi_i = 1, \quad p_{ij} = 0 \tag{72}$$

This system of *n+1* equations with respect to *n* unknowns $\pi_i^{\infty} = 1,2,...n$ has a unique solution.

As an example, consider a two-state case (*n=2):*

$$p_{11}\pi_1^{\infty} + p_{21}\pi_2^{\infty} = \pi_1^{\infty}, \quad p_{12}\pi_1^{\infty} + p_{22}\pi_2^{\infty} = \pi_2^{\infty} \tag{73}$$

Utilizing the constraints in Eqs. (72) one obtains:

$$\pi_1^{\infty} = \frac{1 - p_{22}}{2 - (p_{11} + p_{22})}, \quad \pi_2^{\infty} = \frac{1 - p_{11}}{2 - (p_{11} + p_{22})} \tag{74}$$

Hence on the first sight, there are infinite numbers of unitary matrices u_{ij}, which provide the same ergodic attractor. However, such a redundancy is illusive since the fact that the stochastic matrix p_{ij} has been derived from the unitary matrix u_{ij} imposes a very severe restriction upon p_{ij}:

not only the sum of each row, but also the sum of each column is equal to one, i.e., now in addition to the constrain in Eqs. (72), an additional constraint

$$\sum_{i=1}^{n} p_{ij} = 1 \qquad (75)$$

is imposed upon the stochastic matrix. This makes the matrix p_{ij} doubly stochastic that always leads to an ergodic attractor with uniform distribution of probabilities. Obviously such a property significantly reduces the usefulness of the Quantum recurrent net (QRN). However, as will be shown below, by slight change of the QRN architecture, the restriction (75) can be removed.

12. Multivariate ONR.

In the previous section we have analyzed the simplest quantum neural net whose probabilistic performance was represented by a single-variable stochastic process equivalent to generalized random walk. In this section we will turn to multi-variable stochastic process and start with the two-measurement architecture. Instead of Eq.(60) now we have the following mapping:

$$\frac{1}{\sqrt{2}}\{00...1_{i_1}0...1_{i_2}0...0\} \to \frac{1}{\sqrt{2}}\{00...1_{j_1}0...1_{j_2}0...0\} \qquad (76)$$

i.e., $$I_1 + I_2 \to J_1 + J_2 \qquad (77)$$

where I_1, I_2, J_1 and J_2 are the eigenstates where the unit *1* is at the $i_1^{th}, i_2^{th}, j_1^{th}$ and j_2^{th} places, respectively. Then the transitional probability of the mappings:

$$p_{i_1 i_2}{}^{j_1}(I_1 + I_2 \to J_1) = \frac{1}{2}|U_{j_1 i_1} + U_{j_1 i_2}|^2 \qquad (78)$$

$$p_{i_1 i_2}{}^{j_2}(I_1 + I_2 \to J_2) = \frac{1}{2}|U_{j_1 i_1} + U_{j_1 i_2}|^2 \qquad (79)$$

Since these mapping result from two independent measurements, the joint transitional probability for the mapping (76) is

$$p_{i_1 i_2}{}^{j_1 j_2}(I_1 + I_2 \to J_1 + J_2) = \frac{1}{2}|U_{j_1 i_1} + U_{j_1 i_2}|^2 |U_{j_2 i_1} + U_{j_2 i_2}|^2 \qquad (80)$$

One can verify that

$$\sum_{j=1}^{n} p_{i_1 i_2}{}^{j} = 1, \qquad \sum_{j_1 j_2 = j}^{n} p_{i_1 i_2}{}^{j_1 j_2} = 1, \qquad (81)$$

It should be emphasized that the input patterns *I* interfere, i.e., their probabilities are added according to the quantum laws since they are subjected to a unitary transformation in the quantum device. On the contrary, the output patterns *J* do not interfere because they are obtained as a result of two independent measurements. As mentioned above, Eq. (80) expresses the joint transition probabilities for two stochastic processes

$$I_1 \to J_1 \text{ and } I_2 \to J_2 \qquad (82)$$

which are coupled via the quantum interference. At the same time, each of the stochastic processes (80) considered separately has the transition probabilities following from Eq. (61), and by comparing Eqs. (61) and Eq. (80), one can see the effect of quantum interference for input patterns.

It is interesting to notice that although the probabilities in Eqs. (80) have a tensor structure, strictly speaking they are not tensors. Indeed, if one refers the Hamiltonian *H*, and therefore the unitary matrix *U* to a different coordinate system, the transformations of the probabilities (80) will be different from those required for tensors. Nevertheless, one can still formally apply the chain rule for evolution of transitional probabilities, for instance:

$$p_{i_1 i_2 q_1 q_2}(I_1 + I_2 \to J_1 + J_2 \to Q_1 + Q_2) = p_{i_1 i_2 j_1 j_2} p_{j_1 j_2 q_1 q_2} etc \qquad (83)$$

Eqs. (80) is easily generalized to the case of *l* measurements $l \le n$:

$$p_{i_1 ... i_l q_1 ... q_l} = p_{i_1 ... i_l j_1 ... j_l} p_{j_1 ... j_l q_1 ... q_l} \ etc \qquad (84)$$

$$p_{i_1...i_l j_1...j_l} = \frac{1}{l^l}\prod_{\alpha=1}^{l} |\sum_{\beta=1}^{l} U_{j_\alpha i_\beta}|^2 \tag{85}$$

Now the evolution in physical space, instead of Eq. (51)), is described by the following:

$$a_i(t+\tau) = \sigma_l\{\sum U_{ij}a_j(t)\}, \qquad i = 1.2...n \tag{86}$$

where σ_l is the l-measurements operator.

Obviously, the evolution of the state vector a_i is more "random" than those of Eq. (51) since the corresponding probability distribution depends upon l variables.

Eq. (86) can be included in a system with interference inputs and independent outputs as a generalization of the system (56),(57).

13. QRN with input interference.

In order to remove the restriction (75), let us turn to the architecture shown in Fig. 5 and assume that the result of the measurement, i.e., a unit vector $a_m(t) = \{00...010...0\}$ is combined with an arbitrary complex (interference) vector, Fig. 7.

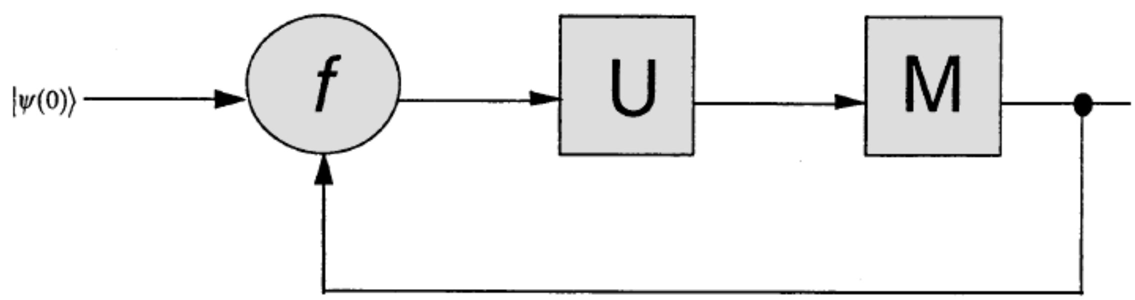

Figure 7. QRN with input interference.

$$m = \{m_1,..m_n\} \tag{87}$$

such that

$$a(t) = [a_m(t) + m]c, \qquad c = \frac{1}{m_1^2 + ...(m_i+1)^2 + ...m_n^2} \tag{88}$$

Then the transition probability matrix becomes

$$p_{ij} = \frac{|U_{j1}m_1 + ...U_{ji}(m_i+1) + ...U_{jn}m_n|^2}{|m_1^2 + ...(m_i+1)^2 + ...m_n^2|} \tag{89}$$

Thus, now the structure of the transition probability matrix p_{ij} can be controlled by the interference vector m.

Eq. (89) is derived for a one-dimensional stochastic process, but its generalization to l-dimensional case is straight-forward.

This architecture produces several interesting algorithms, and one of them is quantum model of emerging grammar, [9]. But in this paper our goal is different: we wish to demonstrate that QRN possesses a distinguished property to violate the second law of thermodynamics and to move from disorder to order without external recourses. That would make the QRN universal in terms of its intelligence capability. In other words, we expect that QRN would implement the model described by Eqs (11) and (12) introduced and analyzed in sections 3,4,5,6, and 7.

14. QRN with nonlinear evolution of probabilities.

So far we were dealing with linear evolution of probabilities (see Eqs. (62) and (71) while evolution of the state vector was always nonlinear(see Eqs. (51),(56) and (57)). Now let us assume that along with the Eq. (51) that is implemented by quantum device, we implement (in a classical way) the associated probability equation (71). At this point these two equations are not coupled yet. Now turning to Eqs. (87), (88), and (89), assume that the role of the interference vector *m* is played by the probability vector π. Then Eqs. (51) and (71) take the form:

$$a_i(t+1) = \sigma_1\{\sum U_{ij}a_j(t)\}, \qquad i = 1.2...n \tag{90}$$

$$\pi_i(t+\tau) = \sum p_{ij}\pi_j(t), \quad i=1,2,...n \tag{91}$$

where $a_i(t) = [\{00...010...0\} + \{\pi_1\pi_2...\pi_n\}]C$ (92)

$$C = \frac{1}{\pi_1{}^2 + ...(\pi_i+1)^2 + ... + \pi_n{}^2} \tag{93}$$

$$p_{ij} = \frac{|U_{j_1}\pi_1 + ...U_{j_i}(\pi_i+1) + ...U_{j_n}\pi_n|^2}{|\pi_1{}^2 + ...(\pi_i+1)^2 + ... + \pi_n{}^2|} \tag{94}$$

and they are coupled. Moreover, the probability evolution (91) becomes nonlinear since the matrix p_{ij} depends upon the probability vector π.

15. Comparison QNR and intelligent particle.

One can associate Eq. (90) with the equation of motion in physical space,(see Eq. (11)) and Eq. (91) – with the Liouville equation describing the evolution of an initial randomness in a probability (virtual) space, (see Eq. (12)). In QNR architecture, Eq. (90) is always nonlinear (due to quantum collapse), while Eq. (91) is linear unless it is coupled with Eq. (90) via the feedback (92). Therefore, one arrives at two fundamentally different dynamical topologies of QRN: the first one is linked to Newtonian physics where equation of motion is never coupled with the corresponding Liouville equation, and the second one can be linked to quantum physics (in the Madelung version of the Schrödinger equation) where the Hamilton-Jacobi equation is coupled with the corresponding Liouville equation by the quantum potential. It is interesting to note that the randomness in Eqs. (11) and (12) is caused by the failure of the Lipchitz conditions accompanied by the blow-up type of instability, (see section 4), while the randomness in QNR is a fundamental quantum phenomenon associated with the quantum collapse as a result of measurement.

Now the following question could be asked: why Eqs.(11) and (12) that describe performance of intelligent particle cannot be simulated directly, and instead, they have to be implemented in the form of QRN prior to simulation? The answer to this question is similar to that given by R. Feynman who explained why quantum phenomena couldn't be simulated with only Newtonian recourses: these phenomena do not belong to the Newtonian world. At the same time, the Schrödinger equation can be computed using a classical computer, and therefore the restriction is imposed only upon simulations, but not upon computations. For a similar reason, Eqs. (11) and (12) cannot be simulated by physical resources without biological parts included since phenomena described by these equations belong neither to Newtonian nor to quantum world: they belong to the world of livings since they violate the second law of thermodynamics. In this context, it is interesting to take a closer look into the architecture of QRN: an initial state, $|\psi(0)>$, is fed into the network, transformed under the action of a unitary operator, *U*, subjected to a measurement indicated by the measurement operator M{ }, and the result of the measurement is used to control the new state fed back into the network at the next iteration. One is free to record, duplicate or even monitor the sequence of measurement outcomes, as they are all merely bits and hence constitute classical information. Moreover, one is free to choose the function used during the reset phase, including the possibility of adding no offset state whatsoever.

As follows from this description, QRN is a hybrid of simulation and computation: the period of action of the unitary operator obviously belongs to quantum simulation. However building a probability vector and feeding it into the net is the element of digital computing. Therefore the hybrid nature of QRN is the reason why QRN could become a universal tool for modeling intelligence. As the last step, we have to prove that QRN can violate the second law of thermodynamics, and that will be the subject of the next section.

16. Spontaneous self-organization.

In this section we will demonstrate a relation of non-linear QRN considered above to a concept of a spontaneous self-organization as a component of life and intelligence. As shown in section 11, a linear QRN eventually approaches an attractor in probability space (see Eq. (74)) that represents a stationary stochastic process, and this attractor does not depends upon initial conditions. Therefore, from the viewpoint of information processing, this attractor performs generalization by placing all possible entry patterns in the same class. Let us ask now the following question: can the system (73) change its evolution, and consequently, its limit distribution, without any external "help"? The formal answer is definitely positive. Indeed, if the transition matrix depends upon the current probability distribution

$$p = p(\pi) \qquad (95)$$

then the evolution (73) becomes nonlinear, and it may have many different scenarios depending upon the initial state π^0. In particular case (71), it can "overcome" the second law of thermodynamics decreasing its final entropy by using only the "internal" resources. The last conclusion illuminates the Schrödinger's statement that 'life is to create order in the disordered environment against the second law of thermodynamics". Indeed, suppose that the selected unitary matrix is

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad (96)$$

Then the corresponding transition probability matrix in Eq. (71), according to Eq. (61) will be doubly-stochastic:

$$p = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \qquad (97)$$

and the stochastic process (71) is already in its thermodynamics limit (97), i.e.,

$$\pi_1 = \pi_2 = 1/2 \qquad (98)$$

Let us assume that the objective of the system is to approach the deterministic state

$$\pi_1 = 1, \quad \pi_2 = 0 \qquad (99)$$

without help from outside. In order to do that, the system should adapt a feedback (95) in the form:

$$a = (a_1, a_2), \qquad a_1 = -2\pi_1, \qquad a_2 = 1 \qquad (100)$$

Then, according to Eqs. (61), the new transition probability matrix p_{ij} will be:

$$p_{11} = \frac{\pi_1^{\ 2}}{2\pi_1^{\ 2} - 2\pi_1 + 1} \qquad p_{12} = \frac{(1-\pi_1)^2}{2\pi_1^{\ 2} - 2\pi_1 + 1}$$

$$p_{21} = \frac{(1+\pi_1)^2}{2\pi_1^{\ 2} + 2} \qquad p_{22} = \frac{(1-\pi_1)^2}{2\pi_1^{\ 2} + 2} \qquad (101)$$

Hence, the evolution of the probability π_1 now can be presented as:

$$\pi_1^{(n+1)} = \pi_1^{(n)} p_{11} + (1 - \pi^{(n)}) p_{21} \qquad (102)$$

in which p_{11} and p_{22} are substituted from Eqs. (101).

It is easily verifiable that

$$\pi_1^{\ \infty} = 1, \ \pi_2^{\ \infty} = 0 \qquad (103)$$

i.e., the objective is achieved due to the "internal" feedback (100).

The application of QRN-based self-organization model to common sense decision-making process has been introduced in [10].

As follows from Eqs.(99) and (103), due to the built-in feedback (100) and without any external effort, the system moved from the state of maximum entropy to the state of minimum entropy, and that violates the second law of thermodynamics. This means that such a system does not belong to physical world: it is neither a Newtonian nor a quantum one. But to what world does it belong? Let us recall again the Schrödinger statement (Schrödinger. 1945): "life is to create order in the disordered environment *against the second law of thermodynamics*". That gives a hint for exploiting the effect of self-organization for modeling some aspects of life, and that makes QRN a universal tool of AI that can compete with human intelligence.

17. Discussion and conclusion.

The objective of this paper is to relate the concept of intelligence to the first principles of physics, and, in particular, to answer the following question: can AI system composed only of physical components compete with a human? The first part of the answer has been addressed in the sections 2 trough7, the second part – in the sections 8 through 16.

The first seven sections introduce and discuss the concept of an intelligent particle. This concept was inspired by the discovery of the Higgs boson and the following from it claim of completeness of the physical picture of our Universe. However the ability to create Life and Mind out of physical matter without any additional entities is still a mystery. The primary objective of this paper is to presents a ***mathematical*** answer to the ancient philosophical question, "How mind is related to matter" in connection with this outstanding accomplishment in physics. The paper is inspired by analysis of the Madelung equation and discovery of the origin of randomness in quantum mechanics, [6]. It turns out that replacement of the quantum potential by the information force, while preserving some quantum properties, introduces fundamental changes in the first and the second laws of thermodynamics, and that leads to a mathematical model that captures behavior of livings. The idea of an intelligent particle has been introduced as a first step of physics of life since it does not include such properties as metabolism and reproduction. Instead it concentrates attention to intelligent behavior. At the same time, by ignoring metabolism and reproduction, we can make the system isolated, and it will be a challenge to show that it still can move from a disorder to the order. It has been demonstrated that the model of intelligent particle belongs neither to Newtonian, nor to quantum mechanics. Its departure from Newtonian mechanics is due to a feedback from the underlying Liouville equation to the equations of motion that represents an additional (internal) information force. Topologically this feedback shifts intelligent particles towards quantum mechanics. However since the information force is different from forces produced by quantum potential, the intelligent particle is not quantum, and it can be identified as quantum-classical hybrid. Therefore intelligent particle dwells in an abstract mathematical world rather than in the physical world, as we know it. This means that intelligent particles, in principle, cannot be composed out of physical particles. It also means that it's behavior can be computed, but not simulated using Newtonian or quantum resources.

The next nine sections introduce a model of quantum recurrent nets for implementation of intelligent particles as a challenge to human intelligence.

There are several advantages that can be expected from quantum implementation of recurrent nets. Firstly, since the dimension of the unitary matrix n can be exponentially larger within the same space had it been implemented by a quantum device, the capacity of quantum neural nets in terms of the number of patterns stored as well as their dimensions can be exponentially larger too.

Secondly, QRN have a new class of attractors representing different stochastic processes, which in terms of associated memory, can store complex behaviors of biological and engineering systems, or in terms of optimization, to minimize a functional whose formulation includes statistical invariants.

The details of ORN performance in learning, optimization, associative memory, as well as in generation of stochastic processes can be found in [7] and [11]. However in this paper, the attention is focused on the most remarkable property of nonlinear QRN that is associated with the spontaneous self-organization as a possible bridge to model intelligent behavior. It is important to emphasize that the architecture of that ORN includes a built-in feedback from the probability evolution to the evolution of the state vector, and that leads to such a non-Newtonian property as transition from a disorder to the order without any external interference. And this property provides the capability of QRN to compete with human intelligence.

References.

1. Schrödinger, E**.** What is life, Cambridge University Press, 1944102. Zak, M., 2013, Information: Forgotten Variable in Physics Models, EJTP 10, No. 28, 199–220,
3. Zak, M., 2013, Particle of life: mathematical abstraction or reality? , Nova publishers, NY.
4. G. Whitham, *Linear and Nonlinear Waves* Wiley-Interscience, New York, 1974.
5. Zak, M., "Terminal Model of Newtonian Dynamics," Int. J. of Theoretical Physics, No.32, 159-190, 1992
6. Zak, M., 2014, The Origin of Randomness in Quantum Mechanics, EJTP 11, No. 31 (2014) 1–16
7. M. Zak, Quantum Neural Nets, International Journal of Theoretical Physics, Vol. 37, No. 2, 1998
8, Barlett,M., 1956, An introduction to stochastic processes, Cambridge University press.
9. Zak M., 2000, "Quantum model of emerging grammars", CHAOS SOL F, 11(14), pp. 2325-2330.
10. Zak, M.,2000, Quantum decision-maker, Information Sciences , 2000, 128, pp. 199-215
11. Zak, M., 1999, Quantum analog computing, Chaos, solitons and fractals,Volume 10, Number 10, September 1999, pp. 1583-1620(38