Quantum model of emerging grammars

Michail Zak
Senior Research Scientist (Emeritus)
Jet Propulsion Laboratory California Institute of Technology
Pasadena, CA 91109, USA.

Abstract.
A special class of quantum recurrent nets (QRNs) simulating Markov chains with absorbing states is introduced. The absorbing states are exploited for pattern recognition: each class of patterns is attracted to a unique absorbing state. Due to quantum interference of patterns, each combination of patterns acquires its own meaning: it is attracted to a certain combination of absorbing states which is di□erent from those of individual attractions. This fundamentally new e□ect can be interpreted as formation of a grammar, i.e., a set of rules assigning certain meaning to di□erent combinations of patterns. It appears that there exists a class of unitary operators in which each member gives rise to a di□erent arti®cial language with associated grammar.

1. Introduction.

One of the oldest and most challenging problems is to understand the process of language formation. In this section we will introduce a model of grammar formation based upon a unique property of QRN: the pattern interference. Let us assume that we store letters of the alphabet in the form of the corresponding stochastic attractors ξ_η. Then if some of these letters, say $\xi_{\eta_1}...\xi_{\eta_l}$., are presented to the QRN simultaneously, their processing will be accompanied by quantum interference in such a way that they will converge to a new attractor, say $\xi_{1,2,..l}$. This new attractor preserves the identities of the letters $\xi_{\eta_1}...\xi_{\eta_l}$, but at the same time, it is not a simple sum of these letters. Moreover, any additonal letter $\xi_{\eta_{l+1}}$ may create a totally different new attractor $\xi_{1,2,..l,l+1}$. Actually this phenomenon is similar to formation of words from letters, sentences from words, etc. In other words, the pattern interference creates a grammar by giving different meaning to different combinations of letters. However, although this grammar is imposed by natural laws of quantum mechanics, it can be changed. Indeed, by changing phases of the components H_{ij} of the Hamiltonian, one changes the way in which the patterns interfere and therefore, the "English" grammar can be transformed into "French" grammar etc.

It should be recalled that the ability to create and understand language is the fundamental property of intelligence that distinguishes human from other livings. At this stage, we do not have any evidence that Nature exploits this particular quantum phenomenon for emerging grammars, but we do not yet observe any alternative ways either. Therefore it is safe to apply QRN for modeling artificial intelligent agents like robots rather than human. In this section, based upon a concept of QRN, a new phenomenological formalism for pattern recognition and grammar formation is described.

2. Emerging grammar formalism.

We will start with a QRN that augmented with a classical measurement and quantum reset operation. The design of the one-dimensional version of this network is shown in Fig.1

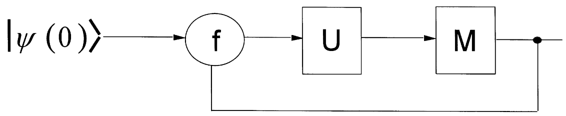

Fig. 1. A one-dimensional quantum recurrent network.

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. 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. For a proof-of-concept, we will start with the following unitary N-dimensional operator

$$U = \begin{pmatrix} U_{11}U_{12}...U_{1n}0...0 \\ U_{12}U_{22}...U_{2n}0...0 \\ .... \\ U_{n1}U_{n2}...U_{nn}0...0 \\ 0..........010...0 \\ .... \\ 0................01 \end{pmatrix} \quad n < N \tag{1}$$

that maps the i^{th} eigenvector into a j^{th} eigenvector

$$\{00...010...0\} \to \{00...010...0\} \tag{2}$$
$$\quad\uparrow_i \qquad\qquad\qquad \uparrow_j$$

with the probability

$$p_{ij} = |U_{ji}|^2 \tag{3}$$

(See Eq. (1.7.12))

Eq. (3) is modified to the following

$$p_{ij} = \frac{|\sum_{k=1}^{n} U_{jk}a_k + U_{ji}|^2}{|\sum_{k\neq i}^{n} a_k{}^2 + (a_i+1)^2|} \tag{4}$$

if each result of the measurement is combined with an arbitrary offset vector

$$|\psi' >= \{a_1 ... a_n\} \tag{5}$$

It should be emphasized that the sum of the output vector in (2) and the offset vector (5) is first calculated, normalized, and then the corresponding quantum re-entering state is prepared.

For the purpose of pattern recognition, the offset vector will be chosen as follows:

$$|\psi_0' >= \begin{cases} \{a_1, a_2, ... a_N\} & if \quad i \le n \\ 0 & if \quad i > n \end{cases} \tag{6}$$

where i is defned by Eq. (2).

Now the probability of the mapping (2) performed by the unitary operator U and the offset vector (6) can be obtained by combining Eqs. (3) and (4), and the transition matrix for the corresponding Markov chain is

$$p = \begin{pmatrix} p_1^{\ 1} \cdots p_1^{\ n} p_1^{\ n+1} \cdots \quad p_1^{\ N} \\ p_2^{\ 1} \cdots p_2^{\ n} p_2^{\ n+1} \cdots \quad p_2^{\ N} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p_n^{\ 1} \cdots p_n^{\ n} p_n^{\ n+1} \cdots\cdots p_n^{\ N} \\ 0\ldots\ldots 0 \quad 10\ldots\ldots\ldots 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 0\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1 \end{pmatrix}, \qquad 0 \le p_i^{\ j} < 1, \qquad \sum_{j=1}^{N} p_i^{\ j} = 1 \qquad \text{(I.2.2.7)}$$

This chain has n transient states $T_q(q = 1.2,\ldots n)$ and N-n absorbing states $A_\gamma(\gamma = n+1, n+2,\ldots N)$ and therefore, regardless of an initial state, the stochastic process eventually will be trapped in one of the absorbing states A_k . However, the probability that it will be a prescribed state A_γ depends upon the initial state. Indeed, as follows from theory of Markov chains, the probability $f_q^{\ k}$ of absorption into A_k from T_q satisfies the system of equations, Feller, W., 1957,

$$f_q^{\ k} = \sum_{j=1}^{n} p_q^{\ j} f_j^{\ k} \quad for \qquad q = 0,1\ldots n; \qquad k = n+1, n+2,\ldots N \qquad (8)$$

Consequently, by appropriate choice of U and $|\psi_0' >$ in Eqs. (1) and (6), one can divide all the initial states into N-n groups such that each state of the group is absorbed (with a sufficiently high probability) into the same prescribed state. Such a performance can be interpreted as pattern classification if each eigenvector introduced to the QRN is associated with the corresponding patterns.

We will not go into more mathematical details here in order to focus attention upon formation of an artificial language instead. For that purpose, suppose that each run of the quantum device is repeated l times while $l \le n$ independent measurements are collected and fed back into QRN. Then, instead of mapping (2), one arrives at the following:

$$\frac{1}{\sqrt{l}}\{0\ldots 0\underset{\uparrow_{i_1}}{1}0\ldots 0\underset{\uparrow_{i_l}}{1}0\ldots 0\} \to \{00\ldots 0\underset{\uparrow_{j_1}}{1}0\ldots 00\underset{\uparrow_{j_m}}{1}0\ldots 0\}, \qquad m \le l \qquad (9)$$

This corresponds to evolution of k different patterns introduced to QRN simultaneously. One can generalize Eq. (4) to the following

$$p_{i_1\ldots i_l j_1\ldots j_m} = \frac{\prod_{\alpha=1}^{m} \left| \sum_{k=1}^{n} U_{j_\alpha k} a_k + \sum_{\beta=1}^{l} U_{j_\alpha i_\beta} \right|^2}{\left| \sum_{k \ne i}^{n} a_k^{\ 2} + \sum_{\beta=1}^{l} (a_{i_\beta} + 1)^2 \right|^m}, \quad m \le l \qquad (10)$$

by considering how each of the recurrent states combined with the offset vector (5) evolves under the action of the unitary operator U . Eq. (10) defines the probability of transition from the set of inputs $i_1...i_l$ to the set of outputs $j_1...j_m$. If $m=l$, and the offset vector is expressed by Eq. (6), the transition probability matrix p_l can be presented in the form similar to p_1 in Eq.(7)

$$p_l = \begin{pmatrix} p_{11...1}{}^{11...1} \cdots p_{11...1}{}^{nn...n} \cdots p_{11...1}{}^{N^l N^l ... N^l} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{nn...n}{}^{11...1} \cdots p_{nn...n}{}^{nn...n} \cdots p_{nn...n}{}^{N^l N^l ... N^l} \\ 0\ldots\ldots\ldots\ldots 010 \ldots\ldots\ldots\ldots\ldots\ldots 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 01 \end{pmatrix}, \qquad (11)$$

This means that the corresponding l-variate stochastic process has n^l transient states $T_q (q = 1,2,...n^l)$ and $N^l - n^l$ absorbing states A_γ, and therefore, $\binom{n}{l}$ combinations of different patterns (in the form of normalized sums of different eigenstates) are mapped onto $N_c \le \binom{n}{l}$ different classes. Hence, the total number of pattern combinations that can be classified by the QRN is

$$S = \sum_{l=1}^{n} \binom{n}{l} = 2^n \qquad (12)$$

Now the performance of the QRN can be given the following interpretation. As soon as the unitary matrix U and the offset vector $|\psi' >$ are chosen (see Eqs. (1) and (6)), all the transition matrices $p_k (k = 1,2,..l)$ are uniquely defined (see Eqs. (4), (7), (10) and (11)). It should be noticed that these matrices do not have to be implemented: they exist in an abstract mathematical space being induced by the operator U and offset vector $|\psi' >$.
If the only one measurement is fed back (l=1), then the transition matrix (7) manipulates by basic patterns-eigenstates that can be identified with ``letters" of an alphabet: by mapping each eigenvector into a corresponding class, it assigns a certain meaning to the letter. If $l$ independent measurements are fed back, $(1 < l \le n)$ then the transition matrix (11) assigns certain meanings to combinations of letters, i.e., to $l$-letter ``words". In order to understand the rules of these assignments, i.e., the ``grammar", let us turn to Eq. (10). As follows from there, in general

$$p_{i_1...i_l}{}^{j_1...j_m} \ne p_{i_1}{}^{j_1} \otimes ... \otimes p_{i_l}{}^{j_m} \qquad (13)$$

i.e., an l-variate stochastic process is not simply the product of l underlying one-dimensional stochastic processes, and the difference

$$\Delta_{i_1...i_l}{}^{j_1...j_m} = | p_{i_1...i_l}{}^{j_1...j_m} - p_{i_1}{}^{j_1} \otimes ... \otimes p_{i_l}{}^{j_m} | \qquad (14)$$

expresses the amount of ``novelty", or new information created by interaction between different patterns via quantum interference. Formally Eq. (14) resembles quantum entanglement that is also responsible for creation of new information; however, actually this entanglement is not quantum: it is a correlation between several classical stochastic processes generated by quantum interference. It should be recalled that classical neural nets where patterns are stored at dynamical attractors, do not have a grammar: any combination of patterns is meaningless unless their storage is specially arranged, and that would require actual implementation of an exponential number of new attractors (see Eq. (12)).

3. Dynamical Complexity.

In this section, we will discuss Shannon and algorithmic complexity of QRN for emerging grammars. Although the concept of complexity is well understood intuitively, its strict definition remains an enigma since there are many different aspects which can be associated with complexity (the number of interacting variables, the degree of instability, the degree of determinism, etc.). Here we will associate dynamical complexity with the degree of unpredictability of the underlying motion. Then the Shannon entropy becomes the most natural measure of dynamical complexity of QRNs:

$$\overline{H} = -\sum_{i=0}^{n-1} \pi_i \log_2 \pi_i \qquad \text{and} \qquad \overline{H}_{\max} \propto \log_2 n \qquad \text{(I.2.3.1) (15)}$$

Let us assume now that the unitary matrix in Eq. (1) is composed of a direct product of *n* 2×2 unitary matrices:

$$U = U_1 \otimes U_2 \otimes \ldots \otimes U_n, \qquad N = 2^n = 2^{q/4} \qquad \text{(I.2.3.2) (16)}$$

where the number of independent components in U_i

$$q = 4n \qquad \text{I.2.3.3 (17)}$$

Then the dynamical complexity of QRN becomes exponentially larger (see Eq. (2):

$$\overline{H}_{\max} \propto \log_2 n \propto n \qquad \text{(I.2.3.4) (18)}$$

although the algorithmic complexity is still expressed by Eq. (17). Thus, QRNs based upon representation (15) generate "complexity" in an exponential rate, and therefore the underlying stochastic processes attain structure of fractals. Indeed, as shown in Shroeder, M, 1991, a continuous version of a Markov process exhibits self-similar structure down to infinitesimal scales of observation. Although the Markov processes generated by QRNs are finite-dimensional, their scales approaches zero exponentially fast when the number of the variables *n* grows only linearly. This means that QRN generate "quantum fractals" which can be applied to image compression, animation, or for a finite-dimensional representation of Weierstrass-type functions which are continuous, but non-differentiable. In contradistinction to classical fractals, quantum fractals are more controllable since their probabilistic structure can be prescribed.

Now suppose that we are interested in generating a stochastic process with prescribed probability distribution. Then the algorithmic complexity becomes important: it will allow us to preserve only $q=4n$ (out of $N = 2^n$) independent characteristics of the distribution (although the stochastic process will be still *N*-dimensional, and its Shannon complexity will be of order of *n*).

The difference between the Shannon and the algorithmic complexities affects the design of the *l* measurements architecture in the following way. Indeed, the input-output relationships require the number of mapping (i.e., quantum circuits) which is polynomial in *N*, i.e., exponential in *n*. However, if the unitary matrix *U* has a direct-product representation (4) then, as follows from the identity:

$$(U_1 \otimes U_2)(a_1 \otimes a_2) = (U_1 a_1) \otimes (U_2 a_2) = Ua \qquad \text{(I.1.2.3.5) (19)}$$

i.e., $a = a_1 \otimes a_2$ (I.2.3.6) (20) and therefore, not only the size of the unitary matrix U and the state vector a, but also the number of mapping circuits for l-measurement architectures become polynomial in n as far as their actual implementation is concerned. In addition to that, in the case (18), n out of l measurements can be performed in parallel.

Eq. (18) is not the only representation of a unitary matrix which preserves its exponential size while utilizing only polynomial resources. Indeed, consider the following combination of matrix products:

$$U = (U_1^{(1)} \otimes ... \otimes U_n^{(1)})(U_1^{(2)} \otimes ... \otimes U_n^{(2)})...(U_1^{(m)} \otimes ... \otimes U_n^{(m)}) \qquad \text{(I.2.3.7)} \quad (21)$$

Here the number of independent components is:

$$q = 4mn \qquad \text{(I.2.3.8)} \quad (22)$$

while the dimensionality

$$N = 2^n = 2^{q/4m} \qquad \text{(I.2.3.9)} \quad (23)$$

In Eq. (23), N and q are associated with the Shannon and the algorithmic complexity, respectively.

Thus, each unitary operator having the structure (15) and supplied with an offset vector of the type (20) generates a new grammar. Since the structure (15) is preserved under matrix products, new operators of the type (23) represent new grammar. In particular, if the time period of each run of the QRN is increased in q times, then the effective unitary operator will be different from the original one and thereby a set of new languages can be generated by the same quantum ``hardware''. In addition to that, Eq. (21) opens up a possibility to build a high-dimensional operator U from low-dimensional components of the same structure. It is worth mentioning that not every language of the possible set of languages is useful. Indeed, the performance of the QRN, and in particular, the assignments of pattern combinations to specific absorbing states is probabilistic. It is reasonable to require that for each selected patterns combination, the corresponding absorbing probability distribution over all possible states has a well-pronounced preference for a certain state; otherwise a word would lose its stable meaning. (It should be noticed that small overlapping of absorbing states is acceptable: it makes the language more colorful by incorporating double-meaning to some words.) As mentioned earlier, stability of the meaning of the basic patterns, i.e., letters, can be achieved by an appropriate choice of the unitary operator (15) and the offset vector based upon solutions of Eq. (22). However, as soon as U and $|\psi_0'>$ are fixed, there is no further control over stability of words' meaning since all the transition matrices p_i are already predetermined. In this situation, one can characterize the effectiveness of the language by the ratio ζ of the number W of useful words to the total number of words S

$$\zeta = \frac{W}{S}, \ S \approx O(2^n) \qquad (24)$$

Hence, in order to maximize ζ, one has to identify such a solution to Eq. (8) which simultaneously stabilizes the meanings of all the letters as well as most of the words. Obviously, in general, this problem is hard.

4. Examples.

In order to demonstrate the existence of effective emerging grammars, consider the following example.

Suppose that in Eqs. (1) and (6) are chosen as follows

$$U = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 & 0 \\ -\sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad a = (0,0,a_3,a_4) \qquad \text{(I.2.4.1)} \quad (25)$$

where φ, a_3, a_4 are real.

Then, applying Eq. (4) one finds the elements of the transition matrix p :

$$p_1{}^1 = p_2{}^2 = \gamma_1 \cos^2\varphi, \qquad p_1{}^2 = p_2{}^1 = \gamma_1 \cos^2\varphi,$$
$$p_1{}^3 = p_2{}^3 = \gamma_1 a_3{}^2, \qquad p_1{}^4 = p_2{}^4 = \gamma_1 a_4{}^2$$
$$p_3{}^1 = p_3{}^2 = p_3{}^4 = p_4{}^1 = p_4{}^2 = p_4{}^3 = 0, \qquad p_3{}^3 = p_4{}^4 = 1 \qquad \text{(I.2.4.2)} \quad (26)$$
$$\gamma_1 = \frac{1}{a_3{}^2 + a_4{}^2 + 1}$$

As follows from Eq. (26), there are two transient states T_1 and T_2, and two absorbing states, A_3 and A_4.

Introducing four input patterns

$$|\psi_1 >= \{1000\}, \quad |\psi_2 >= \{0100\}, |\psi_3 >= \{0010\}, |\psi_4 >= \{0001\} \qquad \text{(I.2.4.4)} \quad (27)$$

as well as their images in the probabilistic space

$$|\pi_1 = \{1000\}, \quad \pi_2 = \{0100\}, \pi_3 = \{0010\}, |\pi_4 = \{0001\} \qquad \text{(I.2.4.5)} \quad (28)$$

first one can write down trivial mapping

$$|\psi_3 > \to \pi_3 \to A_3, \quad f_3{}^3 = 1, \text{and } |\psi_4 > \to \pi_4 \to A_4, \quad f_4{}^4 = 1, \qquad \text{(I.2.4.6)} \quad (29)$$

Other transitions

$$|\psi_1 > \to \pi_1 \to A_3, |\psi_1 > \to \pi_4 \to A_4, \qquad \text{(I.2.4.7)} \quad (30)$$
$$|\psi_2 > \to \pi_2 \to A_3, \text{and } |\psi_2 > \to \pi_2 \to A_4, \qquad \text{(I.2.4.8)} \quad (31)$$

are more complex, and they can be found from Eq. (2.8):

$$f_1{}^2 = p_1{}^1 f_1{}^3 + p_1{}^2 f_2{}^3 + p_1{}^3, \qquad f_2{}^3 = p_2{}^1 f_1{}^3 + p_2{}^2 f_2{}^3 + p_2{}^3, \qquad \text{(I.2.4.9)} \quad (32)$$

whence

$$f_1{}^3 = \frac{a_3{}^2}{a_3{}^2 + a_4{}^2}, \; f_2{}^3 = \frac{a_4{}^2}{a_3{}^2 + a_4{}^2}, \qquad \text{(I.2.4.10)} \quad (33)$$

Similarly one finds

$$f_1{}^4 = f_1{}^3, \qquad f_2{}^4 = f_2{}^3 \qquad \text{(I.2.4.11)} \quad (34)$$

Thus, if

$$a_3 \cong a_4 \qquad \text{(I.2.4.12)} \quad (35)$$

the patterns $|\psi_1 >$ and $|\psi_2 >$ do not have any meaning; with the same probability they can be absorbed by the states A_3 or A_4 . However, if

$$a_3 >> a_4 \text{ or } a_3 << a_4 \qquad \text{(I.2.4.13)} \quad (36)$$

the same patterns are absorbed by only one state A_3, A_4 and that assigns certain meaning to each of them.

For mapping combinations of patterns (27), one has to repeat twice each measurement before feeding it back. Now the input pattern's combinations will be the following:

$$|\psi_{12} >=|\psi_{21} >= \frac{1}{\sqrt{2}}\{1100\}, \qquad |\psi_{13} >=|\psi_{31} >= \frac{1}{\sqrt{2}}\{1010\} \qquad \text{(I.2.4.14)} \quad (37)$$

but their image in the probabilistic space will be di€erent from (4)

$\pi_{12} = \pi_1 \otimes \pi_2 \qquad \pi_{13} = \pi_1 \otimes \pi_3$ (I.2.4.15) (38)

Instead of listing all the 64 elements of the matrix p_2 (see Eqs. (3.11) and (3.12)), we will concentrate upon those that will be used in our analysis. First of all

$p_{ii}{}^{\alpha\beta} = 0\, if \qquad \alpha, \beta \neq i \;\; i = 3,4$

$p_{ii}{}^{\alpha\beta} = 1$ otherwise.

$p_{ij}{}^{\alpha\beta} = 0\, if \qquad \alpha, \beta \neq i \;\; i = 3,4$ (I.2.4.16) (39)

$p_{ij}{}^{\alpha\beta} = 1$ otherwise.

This means that there are four absorbing states: A_{33}, A_{34}, A_{43} and A_{44} ; the rest 12 states (T_{12}, T_{13}, etc.) are transient. Here we will be interested only in the evolution of the pattern's combination $|\psi_{12}>$ (see Eq. (37)) since it is the only one which entangles the patterns $|\psi_1>$and $|\psi_2>$ (see Eq. (27)). (Other combinations: $|\psi_{13}>|\psi_{23}>$etc. are not entangled, and therefore, their evolution can be predicted from the previous analysis as a direct products$|\psi_{13}>\otimes|\psi_{13}>, |\psi_{13}>\otimes|\psi_{13}>$, i.e., it does not have any novelty element.)

Thus, one obtains

$$p_{12}{}^{11} = \gamma_2(\cos\varphi + \sin\varphi)^4, \;\; p_{12}{}^{22} = \gamma_2(\cos\varphi - \sin\varphi)^4,$$

$$p_{12}{}^{12} = \gamma_2(\cos^2\varphi - \sin^2\varphi)^2 = p_{12}{}^{21} \qquad p_{12}{}^{13} = \gamma_2 a_3{}^2(\cos\varphi + \sin\varphi)^2 \qquad \text{(I.2.4.17) (40)}$$

$$p_{12}{}^{14} = \gamma_2 a_4{}^2(\cos\varphi + \sin\varphi)^2 \qquad p_{12}{}^{23} = \gamma_2 a_3{}^2(\cos\varphi - \sin\varphi)^2$$

$$p_{12}{}^{24} = \gamma_2 a_4{}^2(\cos\varphi - \sin\varphi)^2 \qquad p_{12}{}^{34} = p_{12}{}^{43} = \gamma_2 a_3{}^2 a_4{}^2$$

$$p_{12}{}^{33} = \gamma_2 a_3{}^4 \qquad p_{12}{}^{44} = \gamma_2 a_4{}^4$$

where

$$\gamma_2 = \frac{1}{(a_3{}^2 + a_4{}^2 + 2)^2} \qquad \text{(I.2.4.18)} \qquad (41)$$

As follows from the last four equations in (40), there are direct transitions from the pattern combination $|\psi_{12}>$ to the absorbing states. However, in addition to that, there exist many indirect transitions to the same states, for instance, $T_{12} \to T_{13} \to T_{33}, \;\; T_{12} \to T_{14} \to T_{44}$, and these transitions include the entanglement effect that has maxima at $\varphi = \pm\frac{1}{\sqrt{2}}$.

As a result, the pattern combination $|\psi_{12}>$ acquires a new meaning since it cannot be reduced to the direct product of the patterns $|\psi_1>$and $|\psi_2>$.

The performance of this simple QRN becomes more sophisticated if the elements of the unitary matrix *U* and the component of the offset vector *a* in Eq. (15) are complex numbers. Utilizing the properties (21), one can represent a unitary operator *U* in a compressed form gaining exponential dimensionality of *U* with linear resources.

I.2.5. Summary.

Thus, it has been demonstrated that QRN is capable of creating emerging grammars by assigning different meanings to different combinations of letters. The paradigm is based upon quantum interference of patterns, which entangles the corresponding Markov processes, and thereby, creates a new meaning depending upon how different patterns interact. The capacity of the language, i.e., the total number of words in it is exponential in n where n is dimensionality of the basic unitary operator. However, if this operator is presented as a direct product, then the number of words can be made double-exponential in the dimensionality.