2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Relation generation and fast similarity testing for unsupervised learning

Weibo Gong*
Dept. of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01002, USA
gong@ecs.umass.edu

Bo Jiang*
College of Information and Computer Sciences
University of Massachusetts
Amherst, MA 01002, USA
bjiang@cs.umass.edu

Abstract— **Unsupervised learning of concepts is essential for analogical thinking, which is considered a hallmark of human intelligence. We argue that signals in the brain constantly generate relations among their constituents. Unsupervised learning is then accomplished by fast similarity testing of the stored relation sets using resonance transients and time sequenced memory units. We discuss specific algorithms to achieve these functions.**

I. INTRODUCTION

When watching animal documentaries, one realizes that many animals have essentially the same type of abstraction ability as early humans in dealing with the natural environments. The key feature of abstraction is to test similarities of relations among more "concrete" things. Modern humans have the unique advantage of language, which uses words to label concepts and forms an algebra-like system to amplify the abstraction power. Other than that humans and less intelligent animals share many magical functions of the mind, which suggests that the function of abstraction does not necessarily involve a huge number of neurons.

The question of how the mind works is really about the algorithms in the brain [13]. To us this question is specifically on the algorithms for generating relations and testing their similarities, as pursued in [4], [5], [6], [7], [8]. This naturally pushes us to define "relations", and in fact all concepts, in an algorithmic way. Rolf Landauer [12] stressed that "Information is not a disembodied abstract entity; it is always tied to a physical representation. It is represented by engraving on a stone tablet, a spin, a charge, a hole in a punched card, a mark on paper, or some other equivalent." Likewise we would suggest that a relation is a group of neuronal connections and a concept is a group of relations. We note that such groups have no clear-cut boundaries which is a source of the confusions about the nature of a "concept".

In previous works we argued that intelligence should be based on fast retrievable memories of relations among raw sensory signals. In this paper we focus on the generation of such relations. A principle we follow in the development is that at the early stages of relation generation, one should not try to decide which relations would be useful later on. The best strategy for generally applicable intelligence is to gulp every possible relation. Swallow first, learn later. Take vision as an example. The algorithm should try to preserve all information in the visual signals and generate many possible relations among the signal constituents such as pixels and pixel groups. The relations generated would fight for survival over time and those that repeat more or receive more chemical stimulations live better. The "patterns" or "features" should emerge from such evolution processes.

Our approach consists of three major functional procedures. It starts with converting a time-varying matrix of exponentially decaying sinusoidal functions into a single time function that contains rich relational information among the matrix entries. This could be viewed as a high level model for the summation function of some neurons that have multiple inputs and a single output (we emphasize "high level" - see remarks on this later). Once we have a temporal representation of the incoming signal, it will go through a network of filters that picks up and stores sequences of time signal components characterized by the resonant frequencies of the filter components. The learning process starts with storing and clustering such sequences. The basic circuit for storing a time ordered frequency sequence is called a Resonant Chain Unit (RCU) [8].

The above procedures are executed constantly and recursively. The output of a group of RCU clusters could be the time varying matrix being converted into a temporal representation, just like the visual image example discussed above.

To obtain the relations among several spatial-temporal function groups we let them be grasped by one node with multiple input branches and then get the single output to go through the RCU filter network.

We proceed to define the relations among the constituents of the sensory signals. Signal processing textbooks classify sensory signals as spatial or temporal. Roughly speaking, a speech is a temporal signal and a static photo is a spatial one. However, both signals become spatial-temporal as soon as they enter the brain. The audio signals excite the spatially spread cochlear hair cells and the photo pixels acquire time variability via rapid eye movements. Such spatial-temporal signals would aggregate and decompose in order to form relations for many possible combinations of the signal elements. These relations will be strengthened or weakened by the occurrence frequencies as well as some chemicals, forming the experiences in memory.

So the aggregation and decomposition of spatial-temporal signals are of paramount importance in our development. We could start with the aggregation of a finite set of spatial-temporal signals represented by a complex-valued

*This work is supported in part by the U.S. Army Research Office Contract W911NF-12-1-0385.

square matrix, but for conceptual convenience we consider a continuous version instead. We look for a conversion scheme that turns this signal set into a single time function. While this bears structural similarity with a neuron that has multiple inputs and a single output, we will avoid such analogies in order to disentangle from the biological details. In fact we emphasize that our aim is to develop a logically constructed computation system for mind activities and not to mimic the real brain, for the latter contains many legacy codes that are hard to reason about.

II. Spatial-temporal signal conversion

The most obvious case for spatial-temporal signal conversion is in image processing. However the idea is generally applicable for converting a cluster of RCUs to a representing time function. Various attempts of using diffusion to obtain multiscale images, generally referred to as the scale space methods, are based on the property of Gaussian-like diffusion kernels that they do not introduce extraneous features into the diffused image. This is fundamentally due to the fact that Gaussian distribution is a model for the sum of many nearly i.i.d. random components. As such if one views the diffusion on images as spreading the intensity of every pixel around with random jolts for nearly i.i.d directions and strengths then the above property is not surprising. One does not want to use deterministic averaging since the boundary and other parameter choices bring in extraneous features. This has led researchers to conclude that the best blurring is via the Gaussian kernel, which is the Green's function for the diffusion equation. This fact is relevant to our use of diffusion opertors in the signal conversion algorithms.

A ubiquitous signal transmission scheme is to use waves. Human speech signals are waves oscillating in time, which are generated and received by spatially spread oscillating materials. The human speech waves can be easily observed on the iPhone screen when recording. Such waves can be well approximated by a sum of exponentially decaying sinusoid (EDS) signals [1]. One can also imagine that a miniaturization of such a signal could contain just the same amount of information and being used between the communicating neurons in the form of spiking pulses, which can be formed by superposing EDSs. In the following we describe a mechanism for converting the spatial information into a temporal one. A concrete example is to convert a small patch of an image into a time function representation. But the mechanism is more generally applicable (for example to a cluster of neurons).

A famous paper on the relation between the spatial shape information and a temporal function is entitled "Can one hear the shape of a drum?" by Mark Kac in 1966. A follow up work [17] considers a compact Riemannian manifold (M, g) with a smooth closed domain $D \subset M$ and a Laplace operator acting on functions with a Dirichlet boundary condition. Let $p_D(x, y, t)$ be the kernel associated to D, dg the volume form associated to the metric, and

$$\psi(x,t) = \int_D p_D(x,y,t)dg(y) \tag{1}$$

the solution to the initial boundary value problem

$$\frac{1}{2}\Delta_D\psi = \frac{\partial\psi}{\partial t} \quad \text{on} \quad D\times(0,\infty), \tag{2}$$

$$\psi(x,0) = 1 \quad \text{if} \quad x\in D \quad \text{and} \quad 0 \quad \text{if} \quad x\in\partial D. \tag{3}$$

Letting $q(t)$ be the time function representing the heat content of D at time t:

$$q(t) = \int_D \psi(x,t)dg(x), \tag{4}$$

[17] has shown that $q(t)$ admits a small time asymptotic expansion

$$q(t) \sim \sum_{n=0}^{\infty} q_n t^{n/2} \tag{5}$$

where the coefficients q_n are geometric invariants of D.

While the time function $q(t)$ above contains rich information about the "shape" of the domain D, it is not an oscillating time function as desired. This forces us to introduce an imaginary unit i into the equation and write

$$\frac{\partial\psi(x,t)}{\partial t} = \mathrm{i}\frac{1}{2}\Delta_D\psi(x,t). \tag{6}$$

with the domain in an Euclidean space to simplify the illustration. To make the output signal real we could use, among possible schemes,

$$y(t) = \int_D |\psi(x,t)|^2 dx. \tag{7}$$

These considerations motivate us to propose the following dynamic system for the signal conversion:

$$\begin{aligned} -\mathbf{i}\frac{\partial}{\partial t}\psi(x,t) &= \Delta_D\psi(x,t) \\ &+ \int_D G(x,x',t)V(x',t)\psi(x',t)dx'; \quad (8) \\ y(t) &= \int_D \psi(x,t)\psi^*(x,t)dx. \quad (9) \end{aligned}$$

$G(x, x', t)$ could be taken as a Gaussian, with the Dirac delta function as a special case. Some comments are in order.

- While we would eventually turn into numerical algorithms involving only vectors and matrices in $n-$dimensional Euclidean space, the above equation is in the spirit of the Schrodinger equation and deals with vectors and operators in the Hilbert space L_2.
- The complex, time varying $V(x,t)$ represents the spatial-temporal signal to be converted to the time function $y(t)$. As mentioned, pixel-based static images are becoming oscillatory signals in the visual pathway due to the involuntary eye movements, which are better represented by complex valued time signals.
- Comparing to a linear dynamic system in control theory used in previous works the major change is to add the imaginary unit i in the dynamics. This is to generate oscillatory signals which is crucial in communicating information between neurons and neuron clusters. It is also reflecting the reality that the sensory signals are all of the oscillating type. To turn such oscillatory signals

into the spike trains via integrate-and-fire circuits is plausible and natural.

- Recent study [16] indicates that the microsaccades during fixation of the fovea would turn a stationary image patch into a dynamical one in order to maintain the sensitivity of the photo receptors in the retina. This is closely connected to the Laplacian operator in the above equation.

III. All eyes on the same ball

To address the signal conversion and retrieving accuracy issue we use visual accuracy as an example. To reach very high accuracy in visual image processing the system has to use a sort of law of large numbers. This demands many neurons to receive basically iid signals, which in turn demands that such signals are homogeneously spread in a region. How could this be possible? Take a simple gray scale image as an example, how could two neurons located in different places receive the same signal? Using the spectral resolution $\{\lambda_k, \phi_k\}$ of Δ_D, where ϕ_k denotes the eigenfunction corresponding to the eigenvalue λ_k with $\lambda_1 < \lambda_2 < \cdots$, the Dirichlet heat kernel for D can be written as

$$p(x,y,t) = \sum_{k=1}^{\infty} e^{-\lambda_k t}\phi_k(x)\phi_k(y) \tag{10}$$

and the heat content in the domain is

$$\begin{aligned} h(t) &= \int_D\int_D p(x,y,t)dydx \\ &= \sum_{k=1}^{\infty} e^{-t\lambda_k}\left(\int_D \phi_k(x)dx\right)^2. \end{aligned} \tag{11}$$

We note that if the eigenfunction $\phi_k(x)$ is a sinusoid then the integral $\int_D \phi_k(x)dx$ would be related to the phase shift. For example when one expresses a square wave by its Fourier series each sinusoidal base function would be shifted by the phase in the Fourier coefficient. However in boundary value problems the eigenfunctions are alternating but in general not sinusoidal, so the phase shift description is only motivational. Now if we follow the above discussion for inserting an i then we would have the output function as

$$h(t) = \sum_{k=1}^{\infty} e^{-\mathrm{i}t\lambda_k}\left(\int_D \phi_k(x)dx\right)^2. \tag{12}$$

The approximate phase shifts as the integral would be carried by the sinusoidal time functions with different frequencies in the output time function. Since these carriers are orthogonal (or approximately so when a small decay is involved) the output function contains rich information about the "terrain" in the domain D. As we will see later, such output time function also contains rich information about the relations among the parts of the domain terrain.

We consider a 1D case where a function $f(x)$ on an interval $[\alpha,\beta]$ is the "image". The Fourier series coefficients of $f(x)$ form a sequence of complex numbers each with a magnitude and a phase.

Consider a signal sampler that sums the Fourier terms over an interval $[\alpha,\beta]$. We have for the nth Fourier term, $a_n\sin(n\omega x)+b_n\cos(n\omega x) = r_n\sin(n\omega x+\phi_n)$, that (ignoring r_n, which in the conversion scheme (8)(9) would become a factor in the carrier frequency)

$$\begin{aligned} & g^n_{\alpha\beta}(f) \\ =& \int_\alpha^\beta \sin(n\omega x+\phi_n)dx \\ =& \frac{2}{n\omega}\sin\left(\frac{\alpha+\beta}{2}n\omega+\phi_n\right)\sin\left(\frac{\beta-\alpha}{2}n\omega\right) \end{aligned} \tag{13}$$

where all items except ϕ_n are fixed by the sampler position and size in the visual sensory system and are independent of the function $f(x)$. The difference between two functions $f_1(x)$ and $f_2(x)$ will be reflected in their ϕ_n's and eventually in the differences between $g^n_{\alpha\beta}(f_1)$ and $g^n_{\alpha\beta}(f_2)$. Note that all the samplers at different positions are getting the information about the same ϕ_n's. Note also that the phase shifts are much more important than the magnitude as demonstrated in some image processing examples. Furthermore the purpose of such signal processing is not to reconstruct the original image but to be able to tell the similarities and differences of the inputs. The above differences would be accurately estimated and the high resolution perception of the image differences is enabled by "all eyes on the same ball". One may note that the $1/(n\omega)$ factor makes an individual high frequency component small. However the numbers of the high frequency samplers would be larger due to the smaller physical sizes, an observation motivated the method of Mel-frequency cepstrum in audio signal processing.

The above arguments extend to the 2D case for sampling 2D shapes via an individual sampler. In this case one thinks of the image as the superposition of 2D Fourier elements and each 2D sampler is getting the information of the shift amount and orientation for all elements with strong enough presence.

To detect the differences we note that each $g^n_{\alpha\beta}(f)$ is carried by a different frequency. The resonators in the RCU networks described below will be able to receive corresponding quantities. In the brain such quantities are likely to be thresholded into binary to enable or disable certain RCUs. Then the large numbers would compensate for such quantization to achieve accuracy.

IV. Magnus expansion exhibits relational information

The time function $y(t)$ contains rich information about the relations among the constituents of the incoming signal $V(x',t)$. This can be seen from the Magnus expansion described below.

Consider a general time varying linear differential equation for the n-dimensional vector function $Y(t)$

$$\frac{d}{dt}Y(t) = A(t)Y(t), Y(0) = Y_0 \tag{14}$$

with an $n\times n$ matrix $A(t)$. If $[A(t_1), A(t_2)] = A(t_1)A(t_2) - A(t_2)A(t_1) = 0$ for all t_1, t_2 pairs, for example when $A(t) \equiv$

A, we have a matrix exponential solution

$$Y(t) = \exp\left(\int_{t_0}^{t} A(s)\, ds\right) Y_0. \tag{15}$$

In general one could have

$$Y(t) = \exp\left(\Omega(t)\right)\, Y_0, \tag{16}$$

with the series construction

$$\Omega(t) = \sum_{k=1}^{\infty} \Omega_k(t). \tag{17}$$

The Magnus expansion provides such give a solution to the linear time varying matrix equation above, with the first three terms as

$$\begin{aligned}
\Omega_1(t) &= \int_0^t A(t_1)\, dt_1, \\
\Omega_2(t) &= \frac{1}{2}\int_0^t dt_1 \int_0^{t_1} dt_2\ \left[A(t_1), A(t_2)\right], \\
\Omega_3(t) &= \frac{1}{6}\int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 \\
&\quad \left([A(t_1),[A(t_2),A(t_3)]] + [A(t_3),[A(t_2),A(t_1)]]\right).
\end{aligned}$$

Intuitively when $A(t)$ at different ts are commutative the solution is provided by $\Omega_1(t)$. In general the solution needs more terms in the exponent for an exponential representation. The Magnus series does this in a systematic way. Our usage here is simply to support the idea that the spatial-temporal signal transformation via equations (8) and (9) generates an output time function containing rich relational information among different spatial and temporal parts of the input signal $V(x,t)$. When the output time function $y(t)$ hits the RCU networks described later such relational information forms RCU clusters in memory. These relational memory will be used to form common features via repeated experiences, and will serve as the cues and constraints for retrieval.

V. Resonance transients for signal component recognition

To describe the function of an RCU component we consider the recognition of a particular signal component from an input signal composed of many exponentially decaying sine (EDS) functions. The expressing power of such an EDS superposition has been illustrated in literature such as [1][2]. For the current discussion we limit ourselves to the case where all components start at time zero. To treat signals with delayed components one assumes that the delays are large enough to allow the exponential decay to practically separate the EDS blobs.

We analyze the transient behavior of a second order dynamic system stimulated by a decaying sine function input. Specifically we check

$$\begin{aligned}
h(t) &= e^{-\lambda t}\sin(nt) * e^{-\mu t}\sin(mt) \\
&= \int_0^t e^{-\lambda t}\sin(n\tau)e^{-\mu t}\sin(m(t-\tau))d\tau.
\end{aligned} \tag{18}$$

We assume the neural circuits are efficient high Q filters with a very small μ. We also assume the input signal does not decay too fast and thus is with a small λ. The latter can be relaxed. Under these assumptions we have (treating both λ and μ as zero):

$$\begin{aligned}
& h(t) \\
=\ & \frac{m\sin(nt) - n\sin(mt)}{m^2 - n^2} \\
=\ & \frac{m}{m^2-n^2}\left[\sin(nt) - \sin(mt)\right] + \frac{m-n}{m^2-n^2}\sin(mt) \\
=\ & \frac{2m}{m^2-n^2}\left[\sin(\frac{n-m}{2}t)\cos(\frac{n+m}{2}t)\right] \\
+\ & \frac{1}{m+n}\sin(mt).
\end{aligned} \tag{19}$$

When m and n are large the impact of the last term is minimal. The first term is a sine wave at frequency $|n+m|/2$ modulated by a low frequency sine wave at frequency $|n-m|/2$. When $n-m \to 0$ it can be shown that this term goes to t, a basic phenomena in resonance. When n and m are close this term results in an envelop starting going down at around $\pi/|n-m|$. When the difference $|n-m|$ increases this point moves closer to the time origin. Note that the amplitude of the components in the incoming signal does not have much effect on the behavior described above.

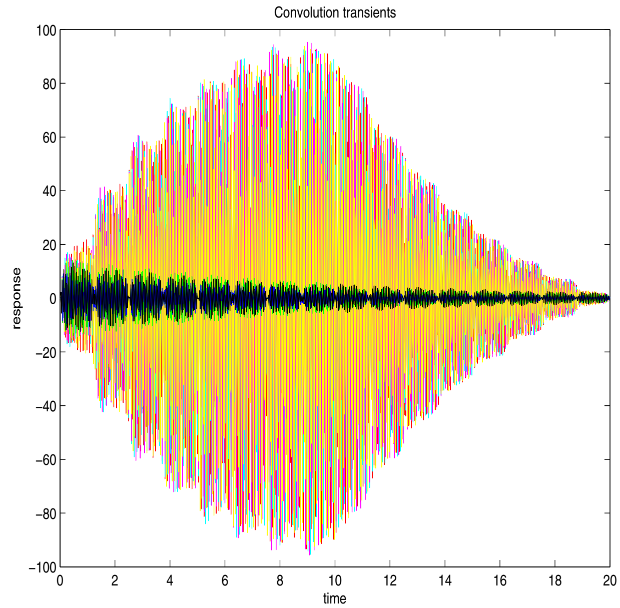

Fig. 1. Using Resonance Transients to Recognize Input Components

Check Figure 1 for a small scale experiment where an EDS signal combination is convolved with single EDS signal components to recognize the signals that are in the combination. The responses for those that are in the input combination are marked by red, cyan, magenta and yellow. The responses for the rest are marked by green, blue and black. Although the frequency differences between the two sets are small one can see the response differences are quite recognizable.

functions and legends of Figure 1. per MATLAB code
t=0.0:0.01:20.0;

```
a=exp(-0.1*t).*sin(500*t);
b=exp(-0.1*t).*sin(505*t);
c=exp(-0.1*t).*sin(510*t);
d=exp(-0.1*t).*sin(515*t);
e=exp(-0.1*t).*sin(495*t);
f=exp(-0.1*t).*sin(490*t);
g=exp(-0.1*t).*sin(520*t);
x=conv(a+b+c+d,a, red;
y=conv(a+b+c+d,b), cyan;
z=conv(a+b+c+d,c), magenta;
w=conv(a+b+c+d,d), yellow;
u=conv(a+b+c+d,e), green;
v=conv(a+b+c+d,f), blue;
s=conv(a+b+c+d,g), black
```

VI. RESONATOR CHAIN AS MEMORY UNIT

We now focus on how to code the EDS signal combinations in memory. To this end we consider a set of resonators sequentially arranged so that each would get excited in the time order in which the matching frequency occurs in the input signal. We call this arrangement a Resonant Chain Unit, or RCU. RCU depends on the resonance mechanism to code information in an unsupervised manner.

An RCU is composed of several resonators with different resonating frequencies. There are fixed time delays between the successive resonators. For an illustrative example consider the case of an RCU consisted of 4 resonators $R1 \sim R4$ with frequencies $f_1 \sim f_4$, respectively. The RCU is build as $R_1 \rightarrow d \rightarrow R_2 \rightarrow d \rightarrow R_3 \rightarrow d \rightarrow R_4$. When R_1 is excited by a matching frequency close to f_1 in the input its output goes through the delay d and gets the R_2 resonator ready. But R_2 has to also receive a signal with a matching frequency close to f_2 to get excited. This process goes on until either all resonators are excited, or the unit excitation is aborted. Figure 2 is a schematic structure of an RCU. RCU could be forced to excite by a signal at the circled box E. Such signals could be due to group connections via simultaneity, attention, and other learning-based connections. RCU output can also branch out (BO in Figure 2) to get other RCU input port ready to respond to an input. RCU input port has a control signal BI which receives signals from other RCU's BO port. The control signals have a binary nature namely they serve as on-off switches.

When an RCU is excited it generates output as the sum of all its resonators with the delays between them. Note that in Figure 2 the sum signal can go out only when R4 is excited. The sum signal goes on to other parts looking for similar RCUs to excite. In the areas immediately behind the signal receptors the RCUs reset to rest condition shortly after the excitation input signal vanishes. However the RCUs in the memory region would sustain the excitation states longer in order to form connections with other excited RCUs to form an RCU cluster. For computer implementation the inter-reset period could be very small.

An RCU could also serve as a node in a super RCU to code more complicated sequences. RCUs may also be used to code spatial-temporal information and to construct hierarchies of memory trees to facilitate fast retrieval of information coded by long RCU sequences. The spatial-temporal signal conversation algorithm is discussed before is generally applicable to convert the signal of a cluster of spatially connected RCUs to a time function.

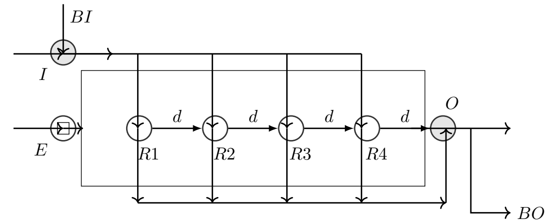

Fig. 2. Basic structure of a Resonator Chain Unit (RCU)

The RCU interactions are quick for the following reasons. When an oscillatory time function hits a group of second order filters its components will selectively excite the resonators with natural frequencies close to the frequency of one of the components, a phenomena similar to the famous Barton's pendulums. While most resonance phenomena takes a while to signify and most textbooks analyze only the steady state behavior, resonance effect in fact accumulates from the moment when the external signal arrives. In other words, when a sinusoid function hits a group of resonators the resonating response curve deviates from others immediately. While the initial differences are small, they are enough for inhibiting other resonators. This is similar to the selection rule in the ordinal optimization method [9][3] referred as horse race rule. As the name implies, often times the transient behavior of a system reveals its potential when ordinal comparison is the decision base. In our current situation it is known in neuroscience that most neurons when excited tend to inhibit other neighboring neurons and render them silent. If we restrict to a mathematically simplified scenario where a causal sinusoid function convolve with the impulse response function of a second order linear system, it can be seen that with the above convolutional horse race selection the input sinusoid function would very quickly select a receiver that resonances with it. This could be the basis for the quickness of image recognition exhibited in human and many animals. In fact this could also be the basis for the quickness of abstract thoughts where relevant concepts in the memory are recalled very quickly.

Resonance transients based quick recall assumes linear signal processing and enables signal aggregation and decomposition in large scale, which are needed to address various binding problems. In previous works [5][6][7][8] we connect the spatial-temporal conversion in linear control theory to the researches relating the geometric shapes to their Laplacian spectrums [11], [15]. Although nonlinear mechanisms are ubiquitous in biological systems including the brain, linear signal processing in these systems could nevertheless be essential in certain signal range.

VII. RCU NETWORKS

We discuss some uses of the RCU networks. Figure 3 shows a small sector of an RCU network to illustrate some connection possibilities.

- **Dictionary style retrieval.** An RCU with BO port can be used to control the access of other RCUs. This is useful in many situations. One important example is to form a dictionary style retrieval. To illustrate this we think of a large dictionary of English words. To check the meaning of a particular word we first code the word into a string of EDSs and then code into a string of RCUs. The output signal of this RCU goes to every input port of the dictionary tree. These input ports are controlled by RCUs that correspond to the first letter of the incoming query word. So only the input port with the matching RCU controller would let in the query word RCU string. The query string drops the first letter RCU code and moves to test the second layer RCU controller with the second letter RCU. And so on. We note again that the control signals serve as on-off switches.
- **Learning with RCU networks** RCUs could form a cluster that would excite together. Excitation of enough number of the member RCUs of a cluster would get the entire cluster excited. The connections among the RCU members in a cluster is through the E port. Such connections would gradually decay and enable learning through repeated use of certain RCU members.
- **Relation recording.**
We consider the playground of mind as a huge fabric in the 3 dimensional space with randomly and densely distributed appositions between the axons and dendrites ready to be connected. The signals running through the links are EDS strings and naturally form RCU clusters for recording and for link control. The links that have been used more often would be easier to access, since such links would have more widely spread access branches than the less used links.
- **Simultaneity recording.** Node $E1, E2, E3$ in Figure 3 denote the output branches that could be sent to other RCU's external triggering point E which, upon receiving enough inputs would excite the RCU. This enables an RCU memory group to be excited together when needed. Different groups may overlap to cover the entire set of simultaneity. Overlapping is used to hook up small groups to form a large group. In visual image coding such overlapping coding is crucial to keep the neighboring information of the image patches. But the mechanism works generally for clusters of RCUs for keeping their connectivity information when multiple concepts interact.

VIII. CONCLUSION REMARKS

Abstraction is the process of sifting the similarities from the instances or the differences among the instances. One of the impressive examples of abstract similarity testing is

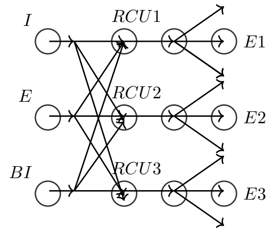

Fig. 3. A small sector of an RCU network

the psychology phenomena referred as analogical reminding, where a sequence of current events remind a possibly remote experience that is only similar in an abstract manner [10][14]. Analogical reminding is featured by the quickness, the abstractness, and the involuntariness of a memory recall carried out by the relatively slow neurons. The spatial/temporal signal conversion and related algorithms described in this paper could help achieving these.

We also note that although the algorithms in this paper heavily rely on transient resonance, in computer algorithms one can easily bypass such physical computations. The purpose is to motivate the development of computer algorithms for unsupervised learning in general artificial intelligence.

REFERENCES

[1] G. Beylkin and L. Monzon. On approximation of functions by exponential sums. *Applied and Computational Harmonic Analysis*, 19(1):17 – 48, 2005.

[2] O. Derrien, R. Badeau, and G. Richard. Parametric audio coding with exponentially damped sinusoids. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7):1489–1501, July 2013.

[3] P. Glynn and S. Juneja. Ordinal optimization: A nonparametric framework. *Proceedings of the 2011 Winter Simulation Conference*, 2011.

[4] W. Gong. Are stochastic dynamics the foundation of intelligence? In *http://www.ieeecss-oll.org/lecture/are-stochastic-dynamics-foundation-intelligence*. 50th IEEE Conference on Decision and Control, 2011.

[5] W. Gong. Can one hear the shape of a concept? In *Proceedings of 31st Chinese Control Conference*, pages 22–26, 2012.

[6] W. Gong. Transient response functions for graph structure addressable memory. In *Proceedings of 52nd IEEE Conference on Decision and Control*, pages 6978–6985, 2013.

[7] W. Gong. Spatial/temporal coding for involuntary concept abstraction using linear dynamic systems. In *Proceedings of 2015 European Control Conference*, 2015.

[8] W. Gong and Y.-C. Ho. On fast retrieval of relational experiences. In *Proceedings of the 2017 American Control Conference*, 2017.

[9] Y.-C. Ho, Q. Zhao, and Q. Jia. *Ordinal Optimization: Soft Optimization for Hard Problems*. Springer, 2007.

[10] D. Hofstadter and E. Sander. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books, 2013.

[11] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73(4, part2):1–23, 1966.

[12] R. Landauer. The physical nature of information. *Physics Letters A*, 217:188 – 193, 1996.

[13] S. Pinker. *How the Mind Works*. Norton, 1997.

[14] S. Pinker. The cognitive niche: Coevolution of intelligence,sociality, and language. *PNAS*, 107(May):8993–8999, 2010.

[15] M. Reuter, F. Wolter, and N. Peinecke. Laplace-beltrami spectra as "shape-dna" of surfaces and solids. *Computer-Aided Design*, 38(4):342–366, 2006.

[16] M. Rucci and J.D. Victor. The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences*, 38 4:195–206, 2015.

[17] M. van den Berg and P. Gilkey. Heat content asymptotics of a riemannian manifold with boundary. *J. Funct. Anal.*, 120:48–71, 1994.