Time, Information and Memory

Kaushik Kumar Majumdar
Systems Science and Informatics Unit, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore
560059, India; e-mail: kmamjumdar@isibang.ac.in

Abstract— **In this work we present a unified mathematical framework, under which time, information and memory are represented. This framework is multidimensional real manifolds. Clock is a system in which a particle traverses the real line (a one dimensional manifold) in uniform velocity. Time is indicated by the position of the particle. Information is represented by the deformation of a multidimensional manifold. This deformation takes some time to take place, during which information is created (encoded) dynamically. When the deformation in the manifold is preserved over time it is called memory. Preservation over short period of time is short term memory and over long period of time is long term memory. Information as deformation of manifold accommodates both Shannon type (information as random occurrence of coded data, devoid of meaning) and semantic type (information as the meaning of data, where the meaning may or may not be unique). Accommodating codons, the unit of genetic information, in this unified framework is very challenging and left as an open problem. If this can be achieved we will get a framework in which it will become possible to study how information encoded in the gene manifests through proteins expression in an organism. Several real life applications of the study have been highlighted.**

Index Terms — **Manifold deformation, memory, semantic information, time, universal clock.**

I. Introduction

Put a scratch mark on a white paper with black ink and it will record the 'memory' of an event, namely, scratch marking the paper with black ink. Information about the event was created by adding a third dimension, called 'color', to the two dimensional sheet of paper and changing the value of the color dimension from zero (no or white color) to nonzero (for black color). Two geometric dimensions of the paper (say X and Y axes) plus the color dimension (say Z axis) make a three dimensional manifold. Introducing change in one or more dimensions information can be encoded into the manifold. If we deform a surface, which is a manifold, its different coordinate values will change, because of its change in shape. This change in shape encodes or creates new information in the form of the new shape. When this shape is static, that is, does not change over time, the shape preserves the information in the form of stored memory as long as the shape remains static. This information is not Shannon information per se,

1

because Shannon information is devoid of any specific meaning [1]. Shannon insisted that attributing an interpretation or meaning to information will destroy its generality and limit its scope. According to him it is only the statistical nature of information that matters. Shannon information theory has remained immensely useful in communication, but has emerged as a kind of bottleneck in biomedical signal processing, where meaning of information contained in the signal is of primary interest to the clinicians and experimentalists [2].

Information with meaning or *semantic information* theory is almost as old as the Shannon information theory. Carnap and Bar-Hillel published their theory of semantic information in 1952 [3]. Unfortunately, there is no consensus on a single, unified definition of semantic information [4]. Here we will not go into the huge debate concerning the philosophy and meaning of information, because at the end of the day they produce little in terms of concrete applications. We will rather take a pragmatic view of information, which is suitable for applications. Information is being created all around us and that is why we can perceive the world. Information is not contained in symbols and signals alone. Even a red rose contains information – information about its existence. Rose is a collection of nested concentric layers of surfaces (petals) with a distinct color, joined together in one end. Mathematically, it is a six dimensional manifold, where the color is an additional three dimensions one for each of the fundamental colors. The smell is yet another dimension (can be more than one dimension also, depending on how the smell is quantified). In order to accommodate a rose in our environment the already existing manifold representing the environment is to be deformed (i.e., changes to be made in three spatial, color and smell dimensions) to make room for the rose. The same is true for a symbol. New information can be created only by deforming the already existing information manifold.

Time is represented by the set of all nonnegative real numbers. If we allow negative time the entire set of real numbers will represent time. Time is measured by a uniform motion machine, such as clock, whose arms rotate at fixed angular velocity. However, the linear distance traversed by a point on an arm of the clock is to be measured on the time axis, that is, on the set of nonnegative real numbers. In other words, time is measured by a uniform motion point (or particle) on the time axis. Information of occurrence of an event can be encoded on the time axis by perturbing the uniform motion point. The point will no longer traverse a flat line, instead it will traverse a signal like path. If we record the signal on a piece of paper like an electrocardiograph or ECG, the information of events (such as heart bits) will get preserved as memory. Here information has been encoded by dynamically deforming (perturbing) a straight line, which is a one dimensional manifold. Memory is created by preserving that deformation in a time invariant form.

The event, information of whose occurrence, has been encoded, may be random or deterministic. If the event is of random occurrence its meaning will be probabilistic. If its occurrence is deterministic, it will have a deterministic meaning, no matter however subjective. In the former case it is Shannon information, in the latter case it is semantic information. A signal can be

partly random, partly deterministic. In that case Shannon and semantic information will be complementary to each other, which is usually the case for biomedical signals [2]. Similar synergistic view of information has already been taken [5].

In the next section we will discuss about time. In section III we will discuss about information. In section IV memory will be discussed. The last section contains concluding remarks.

II. TIME

Time is the greatest synchronizer of all. Time synchronizes the entire universe to behave in a coherent manner. Order and time are synonymous, because the notion of time exists to index classes of events according to the law of mathematical trichotomy. By *mathematical trichotomy* it is meant, given any two real numbers x and y, exactly one of the three (1) $x < y$, (2) $y < x$ and (3) $x = y$ must hold. Now, if A and B are two events, either (i) A starts before B or, (ii) B starts before A or, (iii) A and B starts simultaneously. If starting time of A is indexed by x and starting time of B is indexed by y, then (i), (ii) and (iii) can be substituted by (1), (2) and (3) respectively. Time as ordinal index set for the set of events (that is, there is an order preserving bijective map between set of events and set of time instances) has been well emphasized in scientific literature [6]. The ordered set of real numbers is capable of modeling both temporal and spatial order. *Clock* is the bijective map between the time and the set of real numbers [7], [8] (p. 8), [9] (Appendix A).

Let us consider the mapping $C : \mathfrak{R}^3 \to \mathfrak{R}$, given by

$$x(t) = \cos(t), \ y(t) = \sin(t), \ z(t) = t, \ t \in \mathfrak{R}. \qquad (1)$$

It is easy to verify that the mapping $C(x, y, z)$ is a bijection. Let us take $0 \le t < 2\pi$ and then subdivide the half open interval $[0, 2\pi)$ into 24 equal subintervals of length $\dfrac{\pi}{12}$ each. Now, subdivide each of these subintervals into 60 equal length subintervals of length $\dfrac{\pi}{720}$ each. Again, subdivide each of these subintervals into 60 equal length subintervals of length $\dfrac{\pi}{43200}$ each.

Assuming Sun's relative diurnal motion round the Earth is circular, the angular speed of the Sun is $\dfrac{2\pi}{86400}$ radian/s.

$C(x, y, z)$ given by (1) will be a 24 hour clock if it is equipped with an arm which can make a one single spiral round motion, as shown in Fig. 1, with uniform angular speed $\dfrac{\pi}{43200}$ radian/s. Time is created by uniformly continuous change of position in the clock. It works as the master indexer in the dynamic environment. It can be local, i.e. indexing changes only within a neighborhood but not outside, or it can be global. Here we will remain focused on the global or the universal clock.
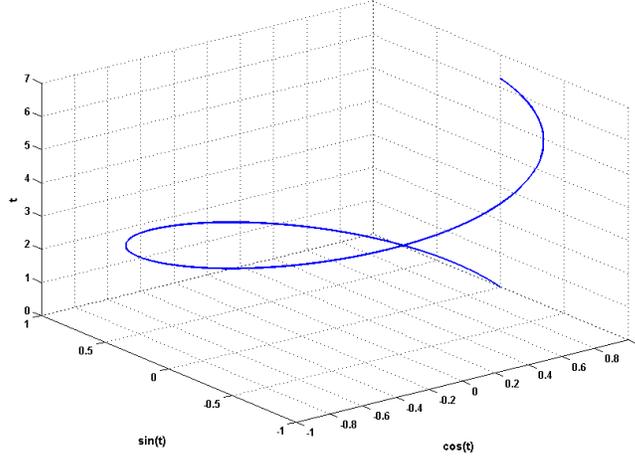

Fig. 1. The clock given by (1) for $0 \le t < 2\pi$, which indicates one day on the Z-axis. 86400 equally spaced points are marked on the spiral trajectory (cannot be seen), the clock arm takes one second to traverse the length between two successive points.

Let us summarize our discussion in this section in the following two definitions:

Definition 1: A *clock* is a differentiable bijection $C : \Re^3 \to \Re$ as defined in (1) with $\dfrac{dC(x(t), y(t), z(t))}{dt} = k$, where t is a parameter and k is a constant.

Definition 2: The parameter t in Definition 1 is *time*. The time we use in particular is for k as the uniform relative speed of the Sun traversing round the Earth.

III. INFORMATION

A. Motivation and Development

It is clear from the previous section that if nothing changes in the environment for ever, there is no significance of time. In a well furnished room at least the arms of the clock are changing position. Time exists, because change exists and each change

4

creates information. Change is indexed by time. So, creation of information is also indexed by time. In other words, temporal information is indexed in time series. This indexing is actually encoding of the temporal information in time series.
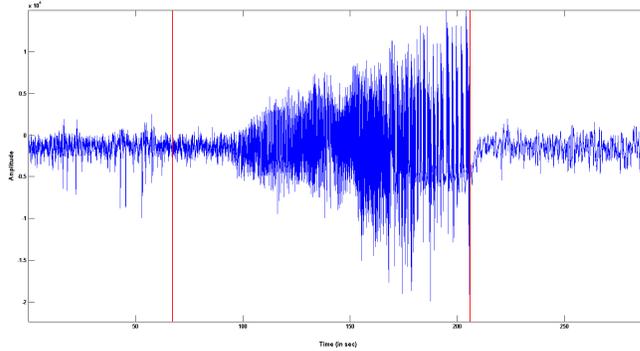

Fig. 2. Time versus amplitude plot of a time series. Here it is brain electrical signal during seizure of an epilepsy patient. Vertical lines indicate start and end of the seizure, which have been determined by visual inspection by an expert.

In Fig. 2 the abscissa is the range of the clock C given by the equations (1), which is time. The time series in Fig. 2 contains some information about the state of brain of an individual, which is encoded along the ordinate. An interesting analogy to observe here is that the time series can be thought of as the trajectory of a particle moving in a force field with one degree of freedom (the freedom is only along the ordinate, the force field is also working along the ordinate). Of course the time series has been taken to be analog or continuous. Assuming the mass of the particle is one, the applied force on the particle at time t is given by $s''(t)$, where $s(t)$ is the time series and s'' denotes the double derivative of s. The work done by the particle in making a ds displacement is $s''(t)ds(t)$. If this takes dt time, the rate at which work is done, i.e., the power of the particle is $P(s(t)) = s''(t)s'(t)$. The work done here is the kinetic energy spent by the particle to give shape to its trajectory. In other words, the shape of the time series $s(t)$ has been given by spending the kinetic energy of the particle with trajectory $s(t)$, which is quite reasonable. But the geometric shape of $s(t)$ contains all the information contained in $s(t)$. This implies information in $s(t)$ is encoded in $s(t)$ or, embedded in $s(t)$ by spending the kinetic energy of the particle. Kinetic energy has transformed into information in the form of the shape of the trajectory or the time series as can be seen in Fig. 2.

Information is encoded into $s(t)$ at time t at the rate at which kinetic energy is dissipated at that instant, which is given by $P(s(t)) = s''(t)s'(t)$, assuming $s(t)$ is double differentiable at t. The time axis in Fig. 2 is the trajectory of a particle with uniform motion, for example, the tip of a clock arm, as was seen in the last section. It does not contain any information ($P(s(t)) = 0$ for all t). But if this straight line is deformed into the time series of Fig. 2 it contains information ($P(s(t)) \neq 0$). Notice, in Fig. 2 both the time axis and the time series are one dimensional surfaces or manifolds. Any of them can be deformed to obtain the other. In other words, deformation can transform a zero information manifold into an information rich manifold and
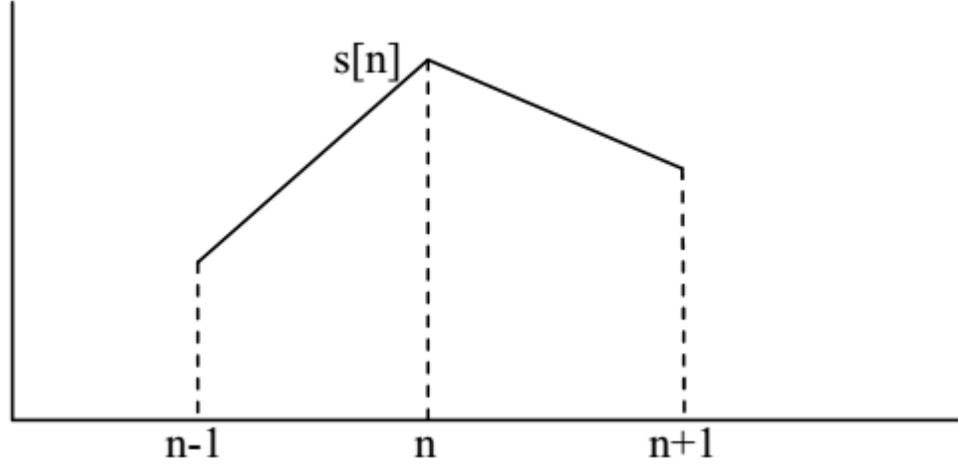
5

vice versa.

Fig. 3. Time (n) versus amplitude (s[n]) plot of the smallest neighborhood of s[n], where n is an interior point. Left product of P(s[n]) is P(s[n-]) = s''[n]*s'[n] and right product of P(s[n]) is P(s[n+]) = s''[n]*s'[n+1], where s'[n] = s[n] – s[n-1] and s''[n] = s'[n+1] – s'[n].

So far $s(t)$ remained analog. If we discretize $s(t)$ the differential operation will become difference operation. Let us replace t by n in the discrete time series $s[n]$. Consider the 3-point neighborhood $n-1$, n and $n+1$ of n, where n is an interior point and $n-1$, $n+1$ are boundary points. *Left product* $P(s[n-])$ and *right product* $P(s[n+])$ of $s[n]$ at n have been defined in Fig. 3.

Theorem 1: Sign change from left to right product at any interior point n for any $s[n]$ takes place in a hierarchical order. That is, if left product is negative, the right product can be negative, zero or positive. Or, if left product is zero, the right product will be zero or positive. Similarly, if left product is positive, the right product can only be positive.

Proof: See Theorem 1 in [10].

This change of sign according to the hierarchical order facilitates determining the shape of the signal or the time series at the three point neighborhood consisting of $n-1$, n and $n+1$ according to the following Theorem 2.

Theorem 2: The 3-point neighborhood or the 3-motif of $s[n]$ on the points $n-1$, n and $n+1$ can have 13 distinct geometric shapes enlisted in Fig. 4.

6

3-motif number	3-motif	LP to RP sign change	s'[n]	s''[n]	s'[n+1]
1			-	+	-
2		- -	+	-	+
3			+	+	+
4		+ +	-	-	-
5			-	+	+
6		- +	+	-	-
7			+	0	+
8			-	0	-
9		0 0	0	0	0
10			0	+	+
11		0 +	0	-	-
12			+	-	0
13		- 0	-	+	0

Fig. 4. Geometric shapes of all the 13 3-motifs that make a discrete time series. How they are being shaped by the force s''[n], momentum on the left (of n) s'[n] and momentum on the right s'[n+1] has been indicated. LP = left product, RP = right product.

Proof: See Theorem 2 in [10].

Summarizing, we can conclude – the model of an analog time series as the trajectory of a particle moving in a force field with one degree of freedom gives us the P-operator, which gives the rate at which information is encoded at any point on the time series. Discretize the time series as well as the P-operator. The sign of left product and right product of P-operator operated on the time series at a point determines the geometric shape of the 3-motif of the time series, of which the point under consideration is an interior point. This shape can have any of the 13 forms shown in Fig. 4.

Note that 3-motif is the smallest motif, which can admit nonlinearity (any two points can be joined by a straight line). It is clear from Fig. 4 how a nonzero $s''[n]$ introduces nonlinearity in the 3-motif, whereas $s''[n] = 0$ preserves linearity. The force $s''[n]$ is molding the shape of the signal. But this very shape contains information contained in the discrete signal or in the discrete time series. In other words, we have proved the following:

Theorem 3: 3-motifs are the smallest elements of information embedded in a digital signal. Alternatively, information is encoded in a digital signal or time series in terms of 3-motifs.

Definition 3: The 3-motifs of Theorem 3 are called *atomic motifs*.

Corollary 1: Peak and trough (motif number 6 and 5 respectively in Fig. 4) are the most information rich 3-motifs out of the 13.

Proof: A peak or a trough is embedded if and only if LP is negative and RP is positive. This can easily be verified by considering the product of signs of first and second difference operators for LP and RP (Lemma 5 in [10]). Here the sign change from LP to RP is a double jump. For all other 3-motifs the sign change is at most with a single jump – negative to zero or, zero to positive (Fig. 4). This completes the proof.

Corollary 1 has wide applications in biomedical signals [11], [12], [13], [14] and financial time series [15] among other signals and time series, where a peak or spike indeed signifies a special event.

B. Shannon Information

According to Shannon's theory of communication *information* is a measure of one's freedom of choice when one selects a message [1] (p. 9). The word information, in this theory, is used in a special sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning. In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information. It is this, undoubtedly, that Shannon means when he says that "the semantic aspects of communication are irrelevant to the engineering aspects." The word information in communication theory relates not so much to what you *do* say, as to what you *could* say [1] (p. 8).

Communication is performed by transmitting and receiving signals containing information. A signal can contain both noise and information. The very words noise and information are relative with respect to the user. Someone looking for 20 to 30 Hz band in a signal and another is looking for 50 to 60 Hz band in the same signal. Only 20 to 30 Hz contains information for one to whom 50 to 60 Hz is noise, whereas to one for whom 50 to 60 Hz is the desired information, 20 to 30 Hz is noise to be filtered out. It is clear that both types of contents (information and noise) of a signal are embedded into its time versus amplitude representation, because this is the representation which uniquely defines the signal in time domain. This representation is also the unique geometric shape that defines the signal. In the previous subsection we have shown how information (both meaningful and meaningless) is embedded into this signal. For example, a peak depending on the circumstances may be a noise peak or a peak signifying a meaningful event. However, in case of meaningful information (semantic information) the shape of the digital signal

expressed by a string of 13 3-motifs, as in Fig. 4, will carry the meaning of the information. In some instances the meaning may be better interpretable in frequency domain or some other transformed domain than in the time domain.

How Shannon information and semantic information complement each other can be elaborated with the following example. A fair coin is tossed once. Before tossing the probability of occurring a head or a tail is ½ each. The Shannon entropy of the event will be 1, assuming the logarithm to have base 2. This means, one binary bit is necessary to contain the information about the probabilistic outcome of the coin toss before the toss (say, 1 for head and 0 for tail). Now, after the toss is over the outcome that has occurred does not involve any more uncertainty. The probability of the outcome is therefore 1. Similarly, the probability of the outcome that has not occurred has the probability of occurring 0. So, the entropy of the event of coin toss after the toss is 0. That is, 0 bit is needed to code the information about its occurrence, or, it does not contain any information. Since there is no uncertainty in it any longer it does not contain any information. But, if we want to encode the face of the coin which has already been tossed we still need one bit to encode that information. Summarizing, we can say, before the toss, Shannon information about the event is 1 and semantic information is 0, whereas after the toss, Shannon information about the event is 0 and semantic information is 1. Whichever way this Shannon or semantic information is encoded it must be encoded through the deformation of a manifold. For example, if it is encoded by writing a binary bit 0 or 1 on a piece of paper, the color dimension of the two dimensional (white) paper surface is altered from white to nonwhite at the coordinate points on the paper which make up the digit 0 or 1. Same will be the case for magnetic surface. The bits will have to be encoded by changing the magnetic polarity at certain location of the surface. Here the information space will consist of surface coordinate(s) plus polarity coordinate.

C. *Generalization*

Now, we are in a position to generalize the notion of information. We propose the following definitions.

Definition 4: An n-dimensional manifold $M_n(\xi_1,....,\xi_n) = \xi_{n+1}$ is *deformed* in the direction of the (n + 1)th coordinate if and only if $M_n(\xi_1 + \Delta\xi_1,....,\xi_n + \Delta\xi_n) - M_n(\xi_1,....,\xi_n) = \Delta\xi_{n+1} \neq 0$

Definition 5: *Information* is created by deformation of manifold. Any deformation of a manifold creates *information*.

Definition 6: An *information space* of n + 1 dimension is an n + 1 dimensional manifold, where an n dimensional manifold is deformed along the (n + 1)th dimension, which is the information dimension.

At this point we must precisely define what we mean by deformation of a manifold. We have shown what it is in one dimension

in the first subsection of this section entitled, "Motivation and development." General manifold deformation will be more complicated to deal with. Therefore, for simplicity, let us remain confined to deformation of surfaces in Euclidean space. First, we will consider a two dimensional surface in $\Re^3$, where $\Re$ is the set of real numbers. Let a two dimensional surface be $f : \Re^2 \to \Re$ is double differentiable at each point. A grey scale image is a good real life example of a two dimensional surface, which contains information. It is pertinent to note that deformation of two dimensional surfaces (such as, by Ricci flow) to represent objects in computers with meaningful shape information is an established paradigm in computer graphics [16].

A grey scale image can be defined as

$$f(x, y) = z \text{ for } a < x < b \text{ and } c < y < d ,$$

$$f(x, y) = 0 \text{ otherwise.} \tag{2}$$

$a, b, c, d \in \Re$ are constant. We can extend the one dimensional model of an analog time series being the trajectory of a particle moving in a force field with one degree of freedom to the two dimensions. Let at each point (x, y) on the Euclidean plane there is a particle moving in a force field with one degree of freedom. The degree of freedom is along the Z-axis. Collection of such particles in the rectangular region $a \le x \le b$ and $c \le y \le d$ will spawn the image surface, where the grey scale value at the point (pixel) (x, y) will be along the Z coordinate. The smooth surface $f(x, y)$ can be imagined to be generated by deforming the xy - plane under the influence of the force along the Z-axis.

Let us parameterize the surface as $f(x(u), y(u)) = z(u)$. f is a free surface, which is free to vary along the Z-axis. Within an infinitesimal neighborhood of the point $(x(u), y(u))$ the deformation of the surface $f(x(u), y(u))$ is along $z(u)$, because the force for deformation is working only along the Z-axis. Since the force is working along the Z-axis, deformation of the surface is taking place along the Z-axis, the information is encoded along the Z-axis at a rate $z''z'$ as shown in III(A). No matter however the morphology of the surface is the information it encodes only in terms of finite geometric shapes as multidimensional extension of the shapes in one dimension, which have already been shown to be finite in discrete as well as in continuous case. We prove the following theorem in several steps through lemmas and theorems.

Theorem 4: In general, information is encoded in terms of finite number of geometric configurations or motifs. At each point on a multidimensional surface there will be one configuration or motif in one coordinate direction. On an n dimensional surface at each point all the n configurations or motifs, one along each coordinate, together will encode the information on the surface at that point.

By *geometric configuration* we mean atomic geometric shape of a continuous time series in an infinitesimal neighborhood of a point. *Motif* is same as configuration, but in discrete case.

Definition 7: A function $f(x_1,.....,x_n) : \mathfrak{R}^n \to \mathfrak{R}$ is *doubly smooth* at a point in $\mathfrak{R}^n$ if and only if $f_{x_i x_j} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ exists at that point $\forall i, j \in \{1,....,n\}$.

Lemma 1: Let $x_{n+1} = f(x_1,.....,x_n)$ be a doubly smooth surface in the n + 1 dimensional information space. Then

$$\frac{d^2 x_{n+1}(u)}{du^2} \frac{dx_{n+1}(u)}{du} = \left(\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \right)^2 f \left(\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \right) f . \qquad (3)$$

Proof: Here we do the proof for two dimensions. Extension to more than two dimensions is straightforward. Let us take the independent variables as x, y and the dependent variable as z. We get $z = f(x,y)$, where the doubly smooth surface f is deformed along the z direction to create information. Let us parameterize the surface. We get

$$z(u) = f(x(u), y(u)) . \qquad (4)$$

$$\frac{dz}{du} = \frac{\partial f}{\partial x} \frac{dx}{du} + \frac{\partial f}{\partial y} \frac{dy}{du} = f_x x' + f_y y' \qquad (5)$$

$$\frac{d^2 z}{du^2} = \frac{\partial f_x}{\partial x} \frac{dx}{du} + f_x \frac{d^2 x}{du^2} + \frac{\partial f_y}{\partial u} \frac{dy}{du} + f_y \frac{d^2 y}{du^2} \qquad (6)$$
$$= f_{xx}(x')^2 + f_{xy} x' y' + f_{yx} y' x' + f_{yy}(y')^2 + f_x x'' + f_y y''$$

Here $f_{xx} = \dfrac{\partial^2 f}{\partial x^2}$, $x' = \dfrac{dx}{du}$, $x'' = \dfrac{d^2 x}{du^2}$, etc.

Notice that the force is acting only along the Z-axis and there is no force along X-axis and Y-axis. The motion of traversing the space along X-axis and Y-axis is uniform (This is akin to the motion along the time axis in the one dimensional case). So, $x'' = 0 = y''$ and $x' = p$, $y' = q$, p and q are constant. p, q are uniform velocities of traversing along X-axis and Y-axis respectively. In one dimensional case time axis was traversed with velocity 1 (which gives the unit of time), We therefore can take $p = 1 = q$. Substituting $x'' = 0 = y''$ and $p = 1 = q$ in (5) and (6) and multiplying them we get

$$z''z' = (f_{xx} + f_{xy} + f_{yx} + f_{yy})(f_x + f_y)$$

$$= \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 f \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) f \qquad . \qquad (7)$$

This completes the proof for two dimensions. It is obvious that the same steps can be followed for dimensions higher than two. This completes the proof.

We have already seen in Section II that time t is nothing but a parameter. In this sense velocity is rate of change of displacement with respect to time. Rate of change of velocity with respect to time is acceleration or force, assuming mass is unity. We can generalize it to any real valued parameter in place of time. In this sense rate at which the surface is deforming with respect to a parameter u, which is algebraically isomorphic and topologically homomorphic to t (actually set of u and set of t are isomorphic and homomorphic to each other respectively), can be taken as the momentum (assuming mass is unity) and the rate of change of momentum with respect to u can be taken as the force.

It is clear from the proof of Lemma 1 that $\dfrac{d^2 x_{n+1}}{du^2} = \left(\sum\limits_{i=1}^{n} \dfrac{\partial}{\partial x_i}\right)^2 f$, which gives the (imaginary) force responsible for deforming the surface $x_{n+1} = f(x_1, \ldots, x_n)$. Therefore, as in the one dimensional case, (3) gives the power of deformation of the surface. In other words, (3) gives the rate at which information is being embedded or encoded in the surface by deforming it.

It is interesting to note that equation (7) gives a straightforward algorithm to calculate the rate of information encoding in an image. When f is an image compute $Df = \left(\dfrac{\partial}{\partial x} + \dfrac{\partial}{\partial y}\right) f$ first. Then compute $\left(\dfrac{\partial}{\partial x} + \dfrac{\partial}{\partial y}\right) Df$ and then multiply both. Rate of information encoding in an image may find good use in content-based image retrieval systems [17].

Theorem 5: A continuous time series is made up of at least 29 different geometric configurations.

Proof: See Theorem 3 in [10].

Lemma 2: A n dimensional doubly smooth surface $f = x_{n+1}$ in $\mathfrak{R}^{n+1}$ can have one of at least 29 different geometric configurations or shapes for the infinitesimal neighborhood of an interior point. In addition, if all the second order partial derivatives of the surface are continuous then there are precisely 7 possible configurations or shapes of the surface for the

infinitesimal neighborhood of an interior point. There will be exactly n configurations at any interior point on f together which will encode the information on f at that point.

Proof: Let us rewrite equation (3) as

$$x_{n+1}^{''} x_{n+1}^{'} = D_n^2 f D_n f , \qquad (8)$$

where $D_n = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}$, etc. u is a parameter, which is algebraically isomorphic and topologically homomorphic to t. So, for all

the operational purposes $x_{n+1}(t)$ can be treated as an ordinary time series, which is double differentiable because the right hand

side of (8) exists and therefore $x_{n+1}(t)$ is continuous.

It   is   clear   from   the   proof   of   Lemma   1 $\frac{d^2 x_{n+1}(u)}{du^2} = \left(\frac{d}{du}\right)^2 x_{n+1} = \left(\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\right)^2 f$ and

$\frac{dx_{n+1}(u)}{du} = \left(\frac{d}{du}\right) x_{n+1} = \left(\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\right) f$, i.e., when f is traversed along x_1 to x_n the deformation of f takes place along
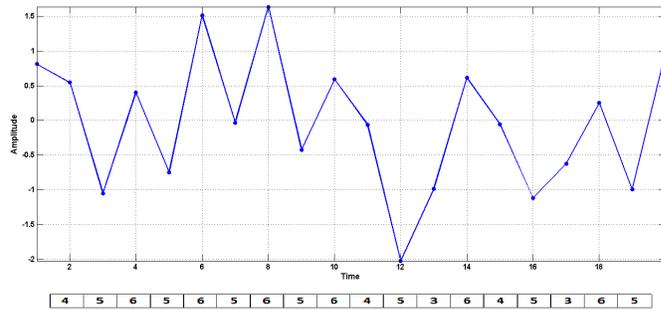
x_{n+1} .

Fig. 5. A small snapshot of a digital signal magnified so that the 3-motif decomposition according to Theorem 2 is clearly visible. The integers in the bottom array are the motif numbers according to Fig. 4.

Imagine the surface f is being traversed along the x_i coordinate for some $i \in \{1,....,n\}$ and its deformation is being

measured along x_{n+1}. $\frac{dx_{n+1}}{du} = \frac{\partial f}{\partial x_i} \frac{dx_i}{du}$, because f is traversed only along x_i. Also $\frac{dx_i}{du} = 1$, and therefore $\frac{dx_{n+1}}{du} = \frac{\partial f}{\partial x_i}$,

$$\frac{d^2 x_{n+1}}{du^2} = \frac{\partial^2 f}{\partial x_i^2} \quad \text{and finally,} \quad \frac{d^2 x_{n+1}}{du^2}\frac{dx_{n+1}}{du} = \frac{\partial^2 f}{\partial x_i^2}\frac{\partial f}{\partial x_i}$$. So, by the same arguments, which were used to prove Theorem 5, the

number of configurations of deforming $x_{n+1} = f$ will be the same as in the one dimensional case and the string made by them

will be oriented along the x_i coordinate. It is illustrated in Fig. 5 for the one dimensional case, where the one dimensional

surface or the line is deformed along the ordinate (say, x_2 coordinate) while the surface is traversed along the abscissa (say, x_1

coordinate) at a uniform rate.

At a particular interior point $p_{n+1} = f(p_1,..., p_n)$, without loss of generality we can take the point

$(0,..., p_i,...0), i \in \{1,....n\}$ as the midpoint of the infinitesimal configuration of f at $(p_1,...,p_i,...,p_n)$ along the x_i

coordinate. The collection of all the n configurations (one each along each coordinate) at $(p_1,...,p_i,...,p_n)$ will give the

information encoded on f by deformation of the surface $x_{n+1} = f$ at the point $(p_1,...,p_i,...,p_n)$.

In addition, if $D_n^2 f$ is continuous (because all second order partial derivatives of f are continuous), x_{n+1}'' is also

continuous. In the infinitesimal neighborhood of a continuously double differentiable point of the time series the double

derivative is either positive or negative or zero. If double derivative is positive, the time series is convex. If it is negative the time

series is concave. If it is zero the time series is a straight line as in Fig. 6.
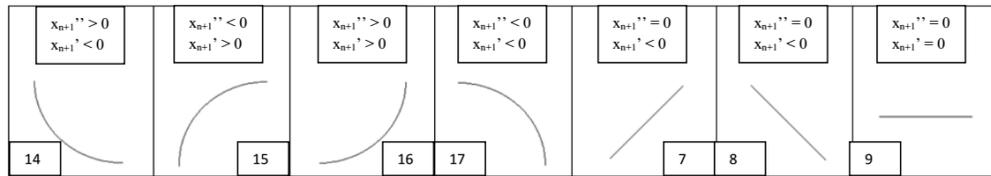

Fig. 6. All the 7 configurations (bottom) those are possible for continuously double differentiable x_{n+1} for different sign combinations of x_{n+1}'' and x_{n+1}'. Sign combinations are shown on the top and configuration numbers at the bottom.

At a continuously double differentiable point x_{n+1} can be convex or concave or a straight line. It can be convex with the first

derivative negative or positive. It can be concave with the first derivative negative or positive. It can be a straight line

with the first derivative negative or positive. Whenever the first derivative is positive x_{n+1} is an increasing function. Whenever

the first derivative is negative x_{n+1} is a decreasing function. Summarizing, at a continuously double differentiable point x_{n+1}

and therefore f too can have one of the 7 configurations as shown in Fig. 6 along each coordinate direction. This completes the

14

proof.

It is to be observed that since $x_{n+1}^{''}(t)$ is continuous in a neighborhood of t if $x_{n+1}^{''}(t)$ is both negative and positive it must

also be 0 at a point. To be more precise, if $x_{n+1}^{''}(t) = 0$ $x_{n+1}^{''}(t+h)$ and $x_{n+1}^{''}(t-h)$ will have different signs for arbitrarily

small $h > 0$. So, in the intervals $(t-h,t)$ and $(t,t+h)$ x_{n+1} can have any of the above mentioned 7 configurations. The

configurations at $(t-h,t)$ and $(t,t+h)$ will have to be interfaced at t in such a manner that $x_{n+1}^{''}(t)$ not only exists but also

remains continuous. This will not add any additional shape other than those shown in Fig. 6.

Lemma 3 (*Discretization lemma*)**:** If the domain of the n dimensional surface function f is discretized then the geometric shape

of the smallest neighborhood of any interior point on the surface is given by m, where m is the Cartesian product of n number

of 3-motifs in Fig. 4. If the interior point on the surface is $p_{n+1} = f(p_1,...,p_n)$, then $(0,...,p_i,...,0), i \in \{1,...,n\}$ is the

interior point or the midpoint of the 3-motif in the i th dimension of m.

Proof: Again we will begin with the two dimensional case, because that will elaborate the multidimensional encoding of

information clearly. Consider equation (4). The XY-plane is discretized in terms of small rectangles, which, borrowing

terminology from image processing, we may call pixel. Each interior (that is, not on the boundary) rectangular pixel has a nine

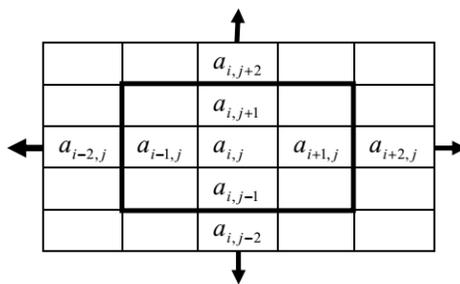pixel neighborhood including itself at the center (Fig. 7).

Fig. 7. A 5 x 5 image matrix, where discretized values of f are shown inside the rectangular pixels (not all pixel values are shown). Thick bordered area is the

nine pixel neighborhood of the pixel at the center, where $f = a_{i,j}$. Arrows indicate the sliding (one pixel at a time) directions of the thick bordered window.

If we follow the convention for discrete partial difference as $\dfrac{\partial f(x,y)}{\partial x} = f(x+1,y) - f(x,y)$ and

15

$$\frac{\partial f(x,y)}{\partial y} = f(x,y+1) - f(x,y), \qquad\qquad \text{then} \qquad\qquad Da_{i,j} = a_{i+1,j} + a_{i,j+1} - 2a_{i,j} \qquad\qquad \text{and}$$

$$D^2 a_{i,j} = a_{i+2,j} + a_{i,j+2} - 4a_{i+1,j} + 2a_{i+1,j+1} - 4a_{i,j+1}$$
$$- 4a_{i,j}$$

Clearly, $D^2 a_{i,j} Da_{i,j}$ will be a unique quantity expressed by $z''z'$. z' and z'' will be calculated when the rectangular window as shown in Fig. 7, is slided by one pixel at a time in one of the coordinate directions shown in Fig. 7 and discrete z is realized as a string of 3-motifs as shown in Fig. 4. Motif representation in both the sliding directions of Fig. 7 will represent the information encoded in the image at each interior pixel. For example, in case of the nine pixel window of Fig. 7, in the horizontal (abscissa) direction it may be the motif no. 6 and in the vertical (ordinate) direction it may be the motif no. 4. The motifs will intersect each other at their mid (interior) point which will be the point on the plane obtained by $f^{-1}(a_{i,j})$. Combination of the horizontal and the vertical motif at each interior pixel will constitute the information encoded by the surface (image) at that point. So, in case of a digital image it is the 13 x 13 alphabet which will constitute the image. Extension to higher dimensions will exactly be the same way. This completes the proof.

Proof of Theorem 4: When f is a doubly smooth surface, by virtue of Lemma 2, at each interior point it encodes information by deformation of the surface in terms of as many number of configurations as the dimension of the surface. By the same Lemma 2 if all the partial double derivatives of f are continuous, only 7 different configurations are possible. If the dimension of the surface is n, at each interior point of f there will be n configurations encoding the information at that point on f, each of which can be any of the 7 types (as shown in Fig. 6).

When f is a discrete surface of n dimension, by virtue of Lemma 3, at each interior point of f information is encoded by deformation of the surface in terms of n motifs. Each of the n motifs can be of 13 different forms as shown in Fig. 4. This completes the proof.

Before closing this section on information, let us make one important observation. Looking at the Fig. 5 it is impossible to make out if the discrete segment is part of a deterministic or random signal, or part of a meaningful or meaningless (noise) signal. It is impossible to determine if it contains Shannon information or semantic (in some sense) information. But in any case whatever information the segment contains is fully encoded in terms of 3-motifs as shown in Fig. 5. The information, irrespective of its type, has been encoded in the signal segment by deformation of the one dimensional surface at each interior point in the

form of any of the 13 3-motifs enlisted in Fig. 4. This is what Definition 5 implies.

IV. MEMORY

A. Physical

Let us look at Fig. 5 once more. It is a snapshot of a time versus amplitude digital signal. Information was being encoded in the signal when it was evolving with respect to time. The snapshot in Fig. 5 is static in time. No fresh information is being encoded in it any more, nor any information content is being removed from it. The information content in the snapshot of Fig. 5 is being stored by preserving the snapshot over the time. In short, the information has become stored memory. The information has been stored as a string of 3-motifs, the atoms of information encoding (see Definition 3).

Some information are time dependent. Some information are not time dependent. In a multidimensional surface $x_{n+1} = f$ general time dependent information are encoded in a framework given by

$$x_{n+1}(t) = f(x_1(t),....,x_{n-1}(t), x_n = t), \qquad (9)$$

where the n th coordinate is the time coordinate and the surface is deformed dynamically with respect to time. If at $t = k$, where k is constant, we want to retrieve the information encoded in the surface by its deformation we will have to rewrite (9) as

$$x_n(u) = f(x_1(u),....,x_{n-1}(u)), \qquad (10)$$

where u is a real parameter that varies at the uniform rate at which the surface $x_n = f$ is traversed along each coordinate axis for retrieving the information encoded in the surface in terms of its deformation. Clearly, at $t = k$ the snapshot of (9) is (10). (10) is called the memory of (9) at time instant k.

Definition 8: *Memory* is information that does not change with respect to time so that it can be retrieved the same way at any point of time. Here 'point of time' can be a value within a finite or infinite interval of $\Re$. (10) is the memory of the information given by (9) at a given time.

By Theorem 4 it is clear that memory can be stored by various combinations of finitely many different configurations or motifs. This 'finitely many' is actually a very small number. The original analog signal in Fig. 5 can be expressed as $x_2(t) = f(x_1(t) = t)$. In the snapshot it was digitized as $x_2(1) = f(t = 1),......,x_2(20) = f(t = 20)$. This was stored as memory, obviously in 3-motif form, out of which the digital signal has been reconstructed in Fig. 5. Processing of this data can be performed at a rate other than the clock rate (offline processing) with the same output as would have been in case of the streaming

17

data. This happens due to preserving the order structure of the streaming data in the stored data.

This order is crucial in determining the type of information the data is containing. If the values $f(1),....,f(20)$ are shuffled randomly the order (which is called temporal, because originally the argument of f was time, but now it can be replaced by any real parameter taking positive integer values) will be destroyed and the data will become random. Alternatively, if only a few values are randomly shuffled and others are left intact, the signal in Fig. 5 will be partially randomized. If the digital signal in Fig. 5 is from a specific location during a specific activity in a human brain. In that sense its information content is semantic information. If it is fully randomized its information content will be Shannon information, and according to Shannon theory its information content will be the highest.

The same mathematical model of memory storage can be extended to computer memories, to archeological and paleontological memories. Computer memories are stored in terms of binary strings. 0, 1 strings can be represented as square wave signals. Square waves can be encoded in terms of 3-motifs numbered 1, 2, 3 and 4 in Fig. 4 with the angles subtended between two straight lines be the right angle.

Archeological structures can be of myriad different architectures. Each architecture can be decomposed into simpler components like dome, pillar, arch, tunnel, staircase, etc. How to represent a dome in terms of configurations and motifs? The dome may not be an entirely differentiable structure, because there may be a spear at its zenith with sharp edges. Also the interface between the dome and the spear may not be differentiable. So, it may be a combination of differentiable and non-differentiable surfaces. At each interior point on the two dimensional continuously doubly smooth surface of the dome the information will be encoded by the first four configurations of Fig. 6, numbered 14, 15, 16 and 17 respectively. One each will be in each of the two coordinate directions. This is because the dome is a convex or a concave structure depending on whether it is seen from above or beneath respectively. The spear head will be the motif number 6 in Fig. 4. The 13 motifs of Fig. 4 and 4 configurations 14, 15, 16 and 17 of Fig. 6 are sufficient for encoding information at any point of most of the real life architectures. Some surfaces are not smooth and will have versatile textures, which will be represented by the 17 motifs or configurations. Color can be represented as three dimensions, each taking values from 0 to 2^7, one for each of the three fundamental colors.

In case of paleontological memories the surface of the fossils will be non-smooth. Again, the texture of the non-smooth surface can be represented by the 17 motifs or configurations, at each point of the surface by two of them one each along a coordinate. The shape and color of the fossils can be represented the way described above. Paleontological information is encoded over surfaces for hundreds of millions of years and becomes memory with virtually no change for over another few millions of years. But this memory is retrieved within a few days or at most in a few years by spatial analysis of shape, texture and color. This is an

18

important observation, where memory takes a long time to form to be retrieved within a short time.

Definition 9: The process of information getting encoded on a manifold to remain invariant over time is known as *learning*.

Of course, learning is the process by which memory is formed. While learning always takes time, memory can be retrieved by traversing the manifold or the surface along all the coordinates at a very fast rate $\left(\dfrac{dx_i(u)}{du} \to \infty, \forall i \right)$, almost instantaneously.

We can call $\dfrac{\dfrac{dx_i(u)}{du}}{\dfrac{dx_i(t)}{dt}}$ as the retrieval / storage ratio or *RS–ratio*, where u is a real parameter and t is time. We have

already taken $x_i^{'}$ to be independent of i. Both in case of archeological and paleontological memory $\dfrac{dx_i(u)}{du}$ will not differ

much, but $\dfrac{dx_i(t)}{dt}$ will differ a lot.

B. Biological

Information and memory are particularly intertwined with each other in the nervous systems of the living organisms. *Semantic memory* in human brain is the memory that is associated with the meaning of words, actions, forms, etc [18]. The learning of words, actions and forms take some time. Some amount of information needs to be provided to the brain to accomplish the learning. After acquiring the required information over a time interval, no matter however long, very little or no further learning takes place in the brain, at least for a prolonged period. Mathematically, we can describe the manifold deformation during learning as

$$x_{n+1}(t) = f(x_1(t),....,x_{n-1}(t), x_n = t \in [a,b]), \qquad (11)$$

where $[a,b]$ is the duration of learning. For $t > b$ the surface given by (11) is not deformed any further, whereas it has been a dynamically deforming surface for $t \in [a,b]$, during which the organism 'learned'. For $t < a$ the information given by (11) was nonexistent. After $t > b$ since $x_{n+1}(t)$ does not change, the expression (11) becomes

$$x_n(u) = f(x_1(u),....,x_{n-1}(u)), \qquad (12)$$

where x_{n+1} has been renamed as x_n and the actual x_n becoming a constant has been discarded. Whatever deformation the surface in (11) underwent in (12) that has been frozen and will not change with respect to time any more. In (12) the shape of the

19

surface can be explored to retrieve the information stored in it in the form of memory. If the memory (12) is semantic the information (11) must also have to be semantic. Memory of a word or action or form may be a collection of a few surfaces like in (12), one each as a snapshot of (11) for a specific value of $t \in [a, b]$. If $a < c_1 < c_2 < b$, for $t = c_1$ (11) becomes

$x_{n+1}(t) = f(x_1(t), \ldots, x_{n-1}(t), x_n = t = c_1)$, which is a surface that no longer changes with respect to t. But it has a shape which varies at all the coordinates from x_1 through x_{n-1}. To trace that shape we replace t by u and x_{n+1} by x_n (the latter is aesthetic, but not essential). Thus we get (12). Similarly, for $t = c_2$. At the time of perceiving

$x_{n+1}(t) = f(x_1(t), \ldots, x_{n-1}(t), x_n = t = c_1)$ and $x_{n+1}(t) = f(x_1(t), \ldots, x_{n-1}(t), x_n = t = c_2)$ brain perceived the first one ahead of the second. At the time of retrieving the memory of the two snapshots how does the brain maintain the same order? A study of place cells in rat hippocampus revealed that when the rat was moving from left to right two particular cells were firing as shown in Fig. 8. When the animal was in Cell 2's place field it fired. When it entered into Cell 1's place field that cell started firing (place fields were overlapped). That is why Cell 2 fired earlier than Cell 1 and their firing patterns in succession represented the sequential nature of traversal by the rat (for more detail see ref. [19]).
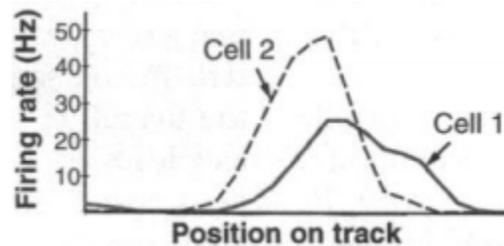

Fig. 8. Mean firing rate of two place cells in rat hippocampus was plotted when the rat was moving from left to right. The firing rate of the cells informed the brain about the sequential traversal of the animal. Taken from [19] with permission.

Interestingly, the same firing sequence by Cell 1 and Cell 2 was replayed during sleep of the animal following the traversal event, but not during sleep preceding the event. This could be due to memory consolidation during sleep [19], which is a well known phenomenon. A sufficiently large number of snapshots for many successive values of t in (11) are preserved in the form of (12), whose replay in the right sequential order informs the animal about the memory of the traversed path. The proposed framework for information and memory allows a generalization of the model from cellular level to neuronal network level.

Population of neurons encodes the activities of the organism better than single cells [2], [20]. As in the study in [19], let the rat hippocampal CA1 neurons be mapped on a two dimensional sheet, on which each cell be located by x-, y-coordinates. Its firing rate is given along z-coordinate. For each time window (say of 200ms duration as taken in [19]) there will be one such three

dimensional snapshot, where the surface $z = f(x, y)$ gives the information of CA1 activity for 200ms during the rat's traversing the path. Successive snapshots during the traversal will map the complete CA1 function during the traversal task. It will be a 3D video, of which firing of Cell 1 and Cell 2 will only be a part. This sequence is consolidated by synaptic potentiation during the sleep. Synapses are like valves, allow information flow only in one direction. So, in a neuronal network as a directed graph, information flow in a particular direction at a particular time interval may preserve the temporal order of the snapshots in the memory.

Although with the currently available recording techniques it is impossible to record from every neuron even in a small region of the brain like CA1, this type of higher dimensional manifold or surface modeling of information and memory has distinct advantages. Nowadays it is possible to record from inside the human brain (while the person is behaving normally) with hundreds of micro-electrodes in the temporal lobe, particularly in hippocampus and adjacent cortical areas. This enables simultaneous recording from tens of single neurons in hippocampus and adjoining areas of behaving humans (predominantly epilepsy patients implanted for clinical evaluation), which have been implicated in learning and memory. Location of those neurons may not precisely be known. Still they can be given an estimated position in the xy-plane, while the firing rate of each cell will be along the z-coordinate. For temporal precision 20 to 200ms [21] time windows may be considered. There will be one 3D snapshot for each time window. Successive windows will give the dynamic deformation profile of the neuronal firing surface. This will certainly be more informative than the single cell studies, which are currently popular among the neuroscientists. In animal studies recording from more number of neurons will be possible and not just by electric potential of single cell but also by optical imaging [22], in which luminosity can be measured along the z-coordinate and information is encoded by deformation of the luminosity surface. Since information is represented by deformation of surface, be it due to electrical firing or intensity of luminosity, both types of information are integrated under one single framework.

V. APPLICATIONS

A. Synchronizability

Synchronization is a ubiquitous phenomenon in nature and society [23]. Synchronization is studied in a network, where each node is an interacting agent and each edge is a connection between two nodes. It is often important to identify nodes which are more responsible for synchronous behavior of the network than the other nodes. The ability of a node to synchronize other nodes with it is called *synchronizability* of that node. An empirical measure of synchronizability of a node has been proposed in [10], the motivation for which is as following.

A node in a network can be regarded as a dynamical system, whose output is the signal that is being generated due to the

underlying dynamics. In a connected network each such system is coupled to all other systems. When several dynamical systems will synchronize with one system in the network (and therefore among themselves as well) that must reflect in their output signals. In other words, the output signals will also have to synchronize, i.e. they will have to be similar to each other in some sense. The scope of random fluctuation in each of them must reduce and they will have to be more 'regular' in some sense. This indicates entropy of each signal should go down. At the same time the 'information power' at the node, which is synchronizing other nodes with itself should go up. We explain this sentence below.

In III(A) we have seen $P(s(t)) = s''(t)s'(t)$, where $P(s(t))$ is the power at which information is being encoded in the signal $s(t)$. $P(s(t))$ is the *information power* of the signal s at time t. The synchronizing node will have to dominate other nodes with its information power in order them to fall in synchrony with itself. This domination will work well for the nodes which are strongly coupled with the dominating node and have similarity in output to some extent with that node. Summarizing, we get a measure of synchronizability for the i th node S_i given by $S_i(s_i(t)) = \dfrac{E(s_i(t))}{P(s_i(t)) * K_i}$, where $E(s_i(t))$ is the

semantic entropy of (digitized) signal $s_i(t)$ emanating out of the i th node and K_i is the total coupling strength of the i th node.

By *semantic entropy* we mean the entropy of the distribution of the 13 3-motifs of Fig. 4. $K_i = \sum\limits_{j=1}^{N} K_{ij}$, where K_{ij} is the

coupling strength between the i th and the j th node, and N is the number of nodes in the network. S_i is to be measured over a time window and $P(s_i(t))$ is the average absolute value of $P(s_i(t))$ over the window. For a digital signal discrete version of P will have to be used. The higher the synchronizability of the i th node the lower will be the value of $S_i(s_i(t))$.

Even E/P gives a good measure of synchronizability, whose efficacy has been tested on human epileptic seizure signals [10]. During seizure across all the nodes (electrodes recording brain electrical signal) E/P goes down, because epileptic seizures induce unusually high synchronization in the brain network. In case of focal seizures, synchronization is most intense in the seizure onset zone (SOZ), where in the implanted recording channels E/P goes much lower than the channels outside this region. It is known that synchronization in the seizure network is dominated by the SOZ. Synchronizabilty may be a useful measure for precisely localizing the SOZ, which is a challenging task and very important from clinical point of view).

B. *Information Flow*

Measuring the flow of information from one node to another in a network is important in the study of network dynamics. Permutation conditional mutual information (PCMI) is a 3-motif based information flow measure [24], where the 3-motifs are

22

3!= 6 permutation motifs. Permutation 3-motifs are subsumed in the 13 3-motifs of Fig. 4, except there are two different types of peaks and two different types of troughs in the permutation 3-motifs. Motifs in Fig. 4 consider two permutation peaks as one single type of peak and same is true for the permutation troughs. The tacit assumption behind the permutation motifs is no two successive values of a digital signal are equal. Motifs in Fig. 4 can represent two successive equal values in a digital signal.

In Fig. 9 the information flow measure between two digital signals has been shown by three different methods namely, Granger causality, PCMI and semantic causality. *Semantic causality* (named after semantic entropy) is a novel measure by the PCMI algorithm [24], but with 13 3-motifs instead of 6 permutation 3-motifs. The three measures do not exactly match with each other. It is clear that towards the end of seizure Y is more dominating than X and information flow is mostly from Y to X, which has been reflected in all the three measures, but most prominently in PCMI and then in semantic causality. On the other hand Granger causality and semantic causality give somewhat similar trend prior to seizure, whereas the trend shown by PCMI is quite one-sided and different.
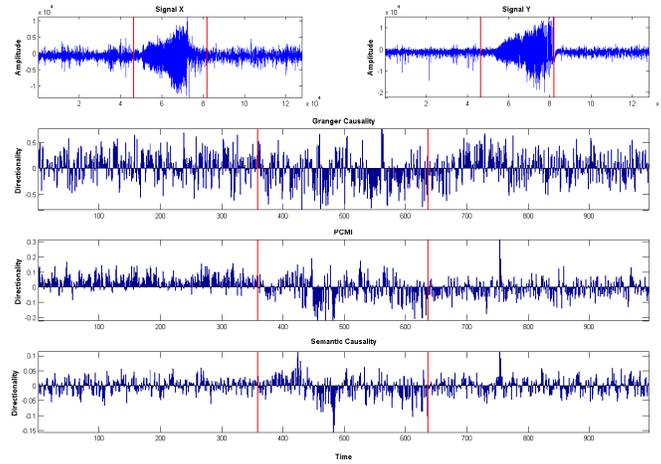

Fig. 9. Signal X and signal Y are two digital intracranial EEG signals of a patient with epilepsy before, during and after a seizure (demarcated by vertical lines indicating start and termination of the seizure). Information flow between the two signals is measured by Granger causality, PCMI and semantic causality (PCMI but with 13 3-point motifs). Directionality above the zero line is from X to Y and below is from Y to X.

C. Computer Vision

From equation (7) it is clear that $\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 f\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) f$ gives the rate at which information is being encoded in the image $f(x,y)$ at a point (x,y). This expression will therefore help to automatically identify points (pixels in digital image, with discrete version of the expression) in an image, where information is changing rapidly. One such situation is grey level areal video image of herds of wild animals moving inside a forest. The contrast between the animals and the surroundings is not much

23

intense. Assuming the image is not contaminated by high frequency noise, $D^2 f\, D f$ will accentuate the contrast between the animals and the surrounding forest. Some values of $D^2 f\, D f$ will keep shifting spatially by small amount in successive still images. Those will represent moving animals. Such techniques will be useful in ecological monitoring.

D. Time Series Data Mining

Time series data mining is an area, where information encoding in terms of motifs has been recognized [25], albeit there is no unique definition of motifs for the time series. The motifs of Fig. 4 are atomic motifs, with which all other motifs can be constructed. There is a transition rule for construction of more complex motifs out of the 3-motifs in Fig. 4 [10], [26]. This type of motif representation will be particularly useful for analyzing micro-array local field potential signals from the nervous systems in order to sort the spikes from single neurons (for detailed review see [27], [28]), which is a challenging data mining task. For general steps of spike sorting and validation algorithms see [29].

E. Codons

Codons are the basic building blocks of genetic information. It is interesting to note that like 3-motifs, the basic building blocks of information contained in a (discrete) time series or a digital signal, codons too are made up of 3 nucleotides. No matter however interesting this analogy is, it is not very helpful to extend our framework to genetic information. Take for example, the RNA codons with all identical nucleotides like AAA, CCC, UUU and GGG. There is only one 3-motif for all 3 values equal, which is a horizontal line (number 9 in Fig. 4), in the scheme proposed in this article. It is not enough to uniquely represent all the four codons with 3 identical nucleotides. Clearly, our framework needs to be extended.

One possible extension is encoding the codons in three dimensions, where each coordinate is expressed as a string of 13 3-motifs of Fig. 4. For RNA codons there will be a 3-motif for A and similarly for C, U and G. Here one challenge will be to come up with motif representation of A, C, U and G in way that they all can be juxtaposed in all possible manners (total 24). In order this to happen the last two points of the 3-motif representing any of the nucleotides will have to be the first two points of any nucleotide including itself. For example, motif 3 in Fig. 4 cannot be followed by motif 1, but can be followed by motif 2. Finding motif representation of the nucleotides maintaining the transition rules [10], [26] will be a combinatorial challenge.

VI. CONCLUSION

In this work we have proposed a common mathematical framework for time, information and memory, which should also include genetic information. However, incorporating genetic information in this framework appears to be technically challenging and is an open problem at this point of time. Possibility of existence of a better framework is also not ruled out. Inclusion of

genetic information and memory in a generalized theory of information encoding, storage and retrieval is imperative for an information theoretic study of evolution. This will be a major thrust for future works in this direction. It should naturally be extended to incorporate the detailed model of circadian clock in the cells [30].

Comparative studies on model generated time series, where we know exactly in which direction information is flowing, need to be performed to establish semantic causality, which appears quite promising (Fig. 9). Additional advantage of semantic causality is that information encoding in a time series in terms of the 13 3-motifs has been mathematically proved [10]. This together with the new synchronizability measure 'entropy divided by the product of information power and coupling strength' [10] can be powerful tools for studying information flow in a complex network leading to synchronization and desynchronization. One advantage of this new measure of synchronizability is that it is node specific, not network specific, like the ones based on graph Laplacian [31].

A general definition of information should accommodate both Shannon's notion and semantic notion of information, which has been presented in Definition 5. Let us consider a signal, which is partly meaningful (say, contains a specific pattern signifying a specific event) and partly meaningless, random (say, contains additive Gaussian noise). The signal contains both semantic and Shannon information. The former part is semantic and the latter is Shannon. The entire signal is created by deformation of a one dimensional manifold or surface. Thus, both Shannon and semantic information can be created (encoded) by deformation of manifold. Therefore, the notion that information is data plus meaning [32] is not necessarily true.

REFERENCES

[1] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois, 1949.

[2] K. K. Majumdar, "Shannon versus semantic information processing in the brain," *Wiley Interdis. Rev. Data Min. Know. Dis.*, e1284, 2018, available at https://doi.org/10.1002/widm.1284.

[3] R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," *MIT Electronics Lab. Tech. Rep. No. 247*, 27 October 1952, available in open source at https://dspace.mit.edu/bitstream/handle/1721.1/4821/RLE-TR-247-03150899.pdf?sequence=1.

[4] L. Floridi, "Is semantic information meaningful data?" *Philos. Phenomenological Res.*, vol. 70(2), pp. 351–370, Mar 2005.

[5] O. Lombardi, F. Holik and L. Vanni, "What is Shannon information?" *Synthese*, vol. 193(7), pp. 1983–2012, 2016.

[6] L. Lamport, "Time, clocks, and the ordering of events in a distributed system," *Commu. ACM*, vol. 21(7), pp. 558–565, 1978.

[7] K. K. Majumdar, "a mathematical analysis of time," *Science Philosophy Interface*, vol. 5, no. 2, pp. 61–69, 2000.

[8] A. T. Winfree, *The Geometry of Biological Time*, 2nd ed., Springer, New York, 2001.

[9] K. K. Majumdar, "Some studies on uncertainty management in dynamical systems using fuzzy techniques with applications," PhD thesis, Indian Statistical Institute, Kolkata, India, Mar 2003.

[10] K. Majumdar and S. Jayachandran, "A geometric analysis of time series leading to information encoding and a new entropy measure," *J. Comp. Appl. Math.*, vol. 328, pp. 469–484, 2018, also available at https://arxiv.org/ftp/arxiv/papers/1810/1810.05900.pdf.

[11] B. S. Todd and D. C. Andrews, "The identification of peaks in physiological signals," *Comp. Biomed. Res.*, vol. 32(4), pp. 322–335, 1999.

[12] N. Acir, I. Oztura, M. Kuntalap, B. Baklan and C. Guzelis, "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural network," *IEEE Trans. Biomed. Eng.*, vol. 52(1), pp. 30–40, 2005.

[13] M. S. Manikandan and K. P. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signals," *Biomed. Sig. Proc. Ctrl.*, vol. 7, pp. 118–128, 2012.

[14] K. K. Majumdar, "Simultaneous multi-channel peaks and troughs in focal epileptic ECoG is more after offset than during the seizure," *Biomed. Sig. Proc. Ctrl.*, vol. 10, pp. 58–64, 2014.

[15] R. Weron and A. Misiorek, "Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models," *Int. J. Forecast.*, vol. 24(4), pp. 744–763, 2008.

[16] M. Jin, J. Kim, F. Luo and X. Gu, "Discrete surface Ricci flow," *IEEE Trans. Visual. Comp. Graphi.*, vol. 14(5), pp. 1030–1043, 2008.

[17] V. N. Gudivada and V. R. Raghavan, "Content-based image retrieval systems," *Computer*, vol. 28(9), pp. 18–22, 1995.

[18] A. Martin and L. L. Chao, "Semantic memory and the brain: structure and processes," *Curr. Opn. Neurobiol.*, vol. 11, pp. 194–201, 2001.

[19] W. E. Skaggs and B. L. McNaughton, "Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience," *Science*, vol. 271, pp. 1870–1873, Mar 1996.

[20] A. P . Gregoropoulos, A. B. Schwartz and R. E. Kettner, "Neuronal population coding of movement direction," *Science*, vol. 233, pp. 1416–1419, Sep 1986.

[21] M. Abeles, "Time is precious," *Science*, vol. 304, pp. 523–524, 2004.

[22] A. Grinvald, R. D. Frostig, E. Lieke and R. Hildesheim, "Optical imaging of neuronal activity," *Physiol. Rev.*, vol. 68, no. 4, pp. 1285–1366, 1988.

[23] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno and C. Zhou, "Synchronization in complex networks," *Phys. Rep.*, vol. 469, no. 3, pp. 93–153, 2008.

[24] X. Li and G. Ouyang, "Estimating coupling direction between neuronal populations with permutation conditional mutual information," *NeuroImage*, vol. 52, no. 2, pp. 497–507, 2010.

[25] A. Mueen, "Time series motif discovery: dimensions and applications," *Wiley Interdis. Rev. Data Min. Know. Dis.*, vol. 4, pp. 152–159, 2014.

[26] K. K. Majumdar and S. Jayachandran, "Semantic information encoding in one dimensional time domain signals," available in open source at https://arxiv.org/ftp/arxiv/papers/1611/1611.01698.pdf, 2016.

[27] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network: Comp. Neural Syst.*, vol. 9, pp. R53–R78, 1998.

[28] E. N. Brown, R. E. Kass and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nat. Neurosci.*, vol. 5, pp. 456–461, 2004.

[29] A. Mitra, A. Pathak and K. Majumdar, "Comparison of feature extraction and dimensionality reduction for single channel extracellular spike sorting," available in open source at https://arxiv.org/pdf/1602.03379.pdf, 2016.

[30] D. B. Forger and C. S. Peskin, "A detailed predictive model of the mammalian circadian clock," *Proc. Nat. Acad. Sci. (USA)*, vol. 100, no. 25, pp. 14806–14811, 2003.

[31] F. M. Atay, T. Biyikoglu and J. Jost, "Synchronization of networks with prescribed degree of distribution," *IEEE Trans. Cir. Syst.–I*, vol. 53, no. 1, pp. 92–98, 2006.

[32] J. C. Mingers, "Information and meaning: foundations for an intersubjective account," *Info. Syst. J.*, vol. 5, pp. 285–306, 1995.