Hippocampal formation mechanism will inspire frame generation for building an artificial general intelligence

Hiroshi Yamakawa

FUJITSU LABORATORIES LTD.
1-1 Kamikodanaka 4-chome, Nakahara-ku, Kawasaki, Kanagawa, 211-8588, Japan

ymkw@jp.fujitsu.com

Abstract The author argues that an artificial general intelligence (AGI) system capable of adapting to various domains autonomously must have the ability to develop domain-specific frames within a practical amount of time; however, current AI technologies are insufficient to achieve this. Frames are knowledge representations which consist of sets of variables. In the frame generation procedure, a significant subprocedure, that of frame candidate generation by variable assimilation, has not yet been realized because of the huge hypothesis space. Representations that can express various relationships among variables in the system can assist in developing this subprocedure, but no such representations have heretofore been known. Through intimate collaboration with neuroscientists, the author searched for clues for such representations in the neuroscience field. Then, the author examined neuroscientific research results to conclude the following: (A) hippocampal formation (HCF) is in charge of frame generation, and (B) distribution equivalent groups (DEGs) are the representations used by HCF for expressing variable relationships. (B) is based on two findings on HCF, namely the phase precession phenomenon and configural association theory. The author used binary-variable assumption to estimate that DEGs exhibit sufficient diversity. Having determined the brain region responsible for a critical function necessary to realize AGI and information representation for that function, this paper offers a foundation for further research into the algorithms used in brain. These results can contribute to the realization of an AGI.

Keywords: computational theory, relationship equivalence, neocortex, variable assimilation, frame problem, hippocampus, relation index, neuroscience

1 Introduction

In general, empirical intelligence is based on comparisons of multiple cases (instance/row) within a "frame." A frame is a well-known declarative knowledge representation; each frame is composed of a frame name and a variable set (including labels for the variables) in which each variable has a value that matches a case. As M. Minsky stated [1], human knowledge is thought to be composed of multiple frames which exist within the brain. As for the current state of artificial intelligence (AI) and machine learning, human beings must design a frame or frame candidates. For example, data analysts use programming languages to define arrays of variables and/or classes as frames. Model selection techniques often chose likely frames from a set of frame

adfa, p. 1, 2011.

candidates (FCs). Tremendous efforts have been made to design various frames in every field; today, domain-specific AI systems can surpass humans in ability.

However, in order to create an artificial general intelligence (AGI), a system must be able to generate domain-specific frames autonomously. Frame generation techniques implemented by selecting a part of a given frame are presently available. For example, feature selection techniques choose useful variables and clustering techniques extract sets of cases as new concepts. Moreover, techniques such as COBWEB [2] and situation decomposition [3] extract concepts by selecting both variables and cases simultaneously. However, these techniques only extract parts of human-designed frames. Thus, we cannot expect creative prediction ability achieved using these techniques to exceed the frame designers' vision. In this paper, I detail a first step toward autonomous frame generation technology beyond this limitation.

2 Computational theory for frame generation

From the above-mentioned background information, it can be surmised that the ability to combine different types of knowledge autonomously is necessary for an AGI. Thus, this section contains an explanation of a computational theory for generating new frames by combining multiple frames together. Here, the equivalence of relationships among variables will play an important role in the practical implementation.

2.1 Frame generation by variable assimilation

There are two kinds of processes to join two frames: the process that assimilates the cases between the frames and the process that assimilates the variables (left side of Fig. 1). In the former process, case assimilations are mediated by the variables shared by both frames ('D' in Fig. 1). Such operations are often used for databases.

By contrast, the latter process generates new frames by joining different cases through assimilating variables (hereinafter, "variable assimilation"). Because this function is related to the frame problem[1], little progress has been seen in research on this subject. The variable assimilation ability is related to high-level intelligences that are well developed in humans, such as analogy, creativity and mimicry. In consideration of this, the heretofore unachieved frame generation ability based on variable assimilation is probably an essential element for creating an AGI. Therefore, a new computational theory to achieve such a function is highly desirable.

On the other hand, some abilities related to variable assimilation have been already realized to a considerable extent because variable assimilation can sometimes be defined as spatial coordinate transformation. For example, many animals can transform from egocentric coordinates to allocentric coordinates for navigation and/or homing. Another example is that visual object recognition systems have the ability to transform input images via operations such as translation, rotation and zoom.

[1] This is a famous fundamental problem for AI; it is caused by the limitless increases in cost for choosing the variables required to perform a target task in an open environment.

2.2 Relationships index supports frame candidate (FC) generation

Because pieces of knowledge in different domains are joined by using a symbolic relationship structure in analogy processes [5], I assumed that combining two subspaces which share some kind of equivalent relationship was likely to be a successful approach for generating a frame by variable assimilation. Here, "subspace" refers to a partial space composed of subsets of variables and "dimensionality" refers to the number of variables in those subsets.

In the example shown in Fig. 1, two subspaces with three variables extracted from two different frames, one containing variables B, A and C and the other containing variables E, G and F, are joined together. Here, the relationship between variables B and A and that between E and G are equivalent, as well as the relationship between A and C and that between G and F. I call the presence of such secondary relationships "relationship equivalence." Based on the relationship equivalence shown in this example, FCs can be generated by assimilating variable B to E, A to G, and C to F.

I denote the total number of variables within a system (including multiple frames) by N and the total number of all available d-dimensional subspaces, each of which contains a unique combination of d variables chosen from all variables, by $S_N(d)$, so $S_N(d) = {}_NC_d$. Thus, the maximum number of paired frames which can be generated is ${}_NC_d \times {}_NC_d$; this limit is a combination of the subspaces. In order to create a practical algorithm against the $O(N^{2d})$ hypothesis space, I needed a technique for drastically reducing the search space.

To focus on the heretofore unachieved computational function, I divided the frame generation procedure into two sequential subprocedures. The first subprocedure, which is for frame candidate (FC) generation, enumerates likely candidates at a practical calculation cost. The second subprocedure, which is for frame verification, selects and improves frames. Because the number of FCs is assumed to be appropriately limited by the former subprocedure, the latter subprocedure can be realized by conventional technology. In the latter subprocedure, criteria such as the mutual information and situation decomposition criteria [3] can be used for evaluating the ap-

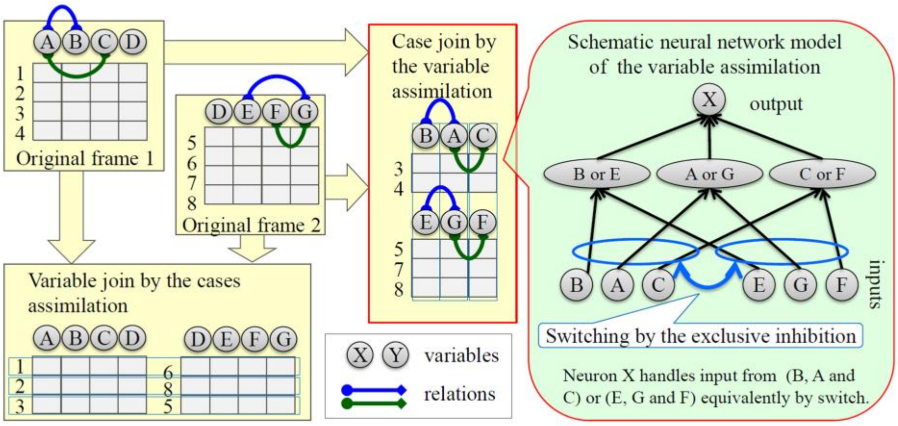

Fig. 1. A variable assimilation (VA) process and its neural network model for frame generation Variables from both frames are matched up with using relational structure in VA process.

propriateness of frames. In the FC generation subprocedure, to enumerate more FCs which can pass the following verification subprocedure, an equivalent relationship among two combined subspaces of an FC should be considered more appropriate as a frame. A relationships index, which categorizes diverse, appropriate subspaces, will help improve the efficiency of FC generation.

This proposed computational theory for frame generation will lead to the development of flexible intelligence technologies and promote the realization of an AGI. The dominant problem is the implementation of an FC generation function because the huge hypothesis space prevents a practical algorithm for implementing this function from being developed. An index of appropriate relationships among variables will assist in the development of an FC generation subprocedure. However, the representation of these relationships is not yet clear.

In the brain, many frame generation abilities are thought to be nonverbal. For example, the spatial coordinate transformation ability (Subsection 2.1) is common to many animals that do not speak any language. It is also known that subconscious intuition and spatial reasoning play important roles in the thinking ability of humans [4]. Research has been conducted on relationship equivalence. The structural mapping theory explains analogies as transferring knowledge from the base field to the target field based on the relationship equivalence among symbolic objects [5]. Neural network models can autonomously discover the relationships among variables from verbal clues [6]. However, these studies are based on symbolic relationships; they cannot explain how to represent nonverbal relationships in the brain.

3 Which brain region handles frame generation?

As mentioned in Subsection 2.1, the brain most likely flexibly generates frames by variable assimilation, but computer systems can presently do little in this regard. Through close collaboration with neuroscientists, such as the fMRI study on human intuition [7], I searched for clues to create the representations and algorithms for such heretofore unachievable functions for computers [8].

3.1 Frames are accumulated on the neocortex and activated

To gather clues about FC generation function from the brain, I desired to identify the brain regions that are responsible for this function. However, since it was difficult to directly specify that region, I instead examined the brain region responsible for accumulating and activating frames. The neocortex is the most likely candidate region for the following four reasons; no other proper candidates have been found.

1. Neural circuits of uniform structure can process general information.
   The neocortex exhibits a uniform structure over a broad area, including the motor, sensory and association areas[2], yet has a general ability to process information in

[2] The neocortex has a functionally differentiated six-layer structure and a column structure. Its basic architecture is the same in all mammals.

diverse ways depending on the domain and modality. Therefore, this region appears suited to accumulating and activating the frames obtained from experience.

2. Parallel switches are required for neural circuits to process frames.
   As shown on the right-hand side of Fig. 1, a schematic neural network that can generate frames necessitates a network that allows for switching in parallel among input variables[3] from different inputs. Support for the possibility that the neocortex is using frames can be found in the fact that it is capable of creating neural networks that switch inputs to local circuits by top-down attention [9].
3. The neocortex is a brain region that is particularly well developed in humans. Mammals, especially humans, have particularly well-developed general intelligence. Therefore, there is a high possibility that the neocortex, which is also particularly well developed in humans, supports general intelligence.
4. The information integration function of the neocortex is also suitable for frames.
   The neocortex is thought to have evolved to provide a multi-modal information integration function that brings together auditory, olfactory and tactile input, which is required for mammals to accurately identify external objects in the dark. The function to select the relevant variables required for reasoning through the use of frames is essential for achieving effective integration of information.

3.2 Hippocampal formation (HCF) works as the frame generator

If the neocortex is the brain region that accumulates and activates frames, it is natural to consider the possibility that frames are also generated within the neocortex. However, I contend that the neocortex is incapable of generating FCs for the reasons to follow in this section. Instead, I believe that hippocampal formation (HCF) is the most probable candidate responsible for generating frames.

Frame generation requires a function to globally integrate the distributed knowledge accumulated locally in individual areas of the neocortex. As already mentioned in Section 2, FC generation requires a function for comparing and categorizing the relationships among variables. If the neocortex has such a function, there must be suitable representations sufficiently diverse to categorize at least 4×10^{12} relationships across different areas of the neocortex[4]. Firing synchronization of neuron populations is a known mechanism for transmitting information directly across different areas of the neocortex; however, the amount of information that can be transmitted by this mechanism is far too small to express the relationships among microscopic neural activities. Therefore, the neocortex lacks representations sufficient to categorize neural activity relationships across different areas, so it cannot generate FCs by itself through such representations.

---

[3] Variables can potentially be associated to various substances such as neurons, minicolumns, hyper-columns and cell assemblies in neural networks within the brain.

[4] Even if each variable is assumed to be a hyper column that is a large unit in the neocortex, there are about 2×10^6 variables in the human neocortex [12]. Thus, the number of relationships between two variables numbers 4×10^{12}.

If the neocortex has no function for generating frames by assimilating the variables dispersed through relationship equivalence, HCF is the only plausible candidate capable of having such a function for the following four reasons:

1. HCFs receive projections from a wide area of the neocortex.
   Because HCFs receive information from a wide area of the neocortex through the entorhinal cortex (EC) [10], they can compare distributed variables on the neocortex to generate frames.
2. HCFs memorize experiences before the neocortex.
   First, memories obtained from the external environment as experience are stored in HCFs. Later, they are gradually transported to the neocortex over a period of several months (in the case of humans) [11]. Therefore, the frame representations in the neocortex are likely to be generated in HCFs.
3. HCFs act together with the frame-executable circuit.
   Because HCFs act in collaboration with the EC they presumably can execute a frame function like the neocortex. Therefore, HCFs can use memories seamlessly while transporting such memories to the neocortex.
4. HCFs contain a local circuit for signal exchange.
   In the subregion of HCFs (dentate gyrus and CA3), single-layer neurons form a circuit-like crossbar matrix [13]. These circuits are suited for switching global inputs and outputs and are probably appropriate for searching for FCs.

Then, HCFs are thought to be responsible for generating various pieces of knowledge as frames and for using these frames in collaboration with ECs. In addition, frames are transported to the neocortex over time. These assumptions are consistent with one explanation of the function of HCFs, the relational theory [14].

4 Distribution equivalent groups (DEGs) represent relationships

In the above arguments, I explained that in order to realize an AGI, a computational theory for generating frames by variable assimilation is a promising approach. I also contended that the neural circuits of HCFs and the ECs provide clues for designing the representations and algorithms necessary for this theory. The most serious problem is how to deal with the huge hypothesis spaces in the process of generating FCs. To overcome this obstacle, the approach of developing an algorithm which uses the relationships among variables is promising. Using clues from findings on HCFs, in this section I discuss probable candidates for representations that express the relationships among variables.

4.1 DEGs as particular representations for HCFs

The long history of research on HCFs provides an ample amount of valuable experimental findings as well as theoretical hypotheses. The following two HCF-specific findings are significant in restricting representation of relationships.

First, there is the theta phase precession phenomenon, in which an approximately 5 Hz theta rhythm is generated in HCFs of an animal moving around an experimental field and discrete case sequences, which contain from seven to twelve cases, appear in the theta phase in a time-compressed manner [15]. Secondly, according to the configural association theory [16], which explains the functional role of HCFs, HCFs associate combinations (complexes) of stimuli rather than individual stimuli with the meaning of behaviors. Although an animal suffering from lesions on its HCFs can form simple associative memories between stimuli, it cannot perform tasks by memorizing a number of stimuli all together (e.g., a transverse patterning problem). These particular characteristics of HCFs lead me to assume that each relationship representation is likely to be a set of about ten cases in a multi-dimensional subspace.

Here, I define a Distribution Equivalent Group (DEG), which consists of about ten cases set in a multidimensional subspace in consideration of variable exchange symmetry. This symmetry means that if two distributions are identical when exchanging variables within the subspace, then they belong to same DEG. After all, DEGs are thought to be the representation used by HCFs for variable relationships.

4.2 Binary variable DEGs for estimation

In order to estimate the number of DEGs, I assumed a binary variable {0, 1} and composed lattices as a set of cells as shown in Fig. 2-A. Cells which contain cases are shaded red. All binary lattice patterns for low dimensional (d = 1 or 2) DEGs are shown in Fig. 2-B. Here, the number of cells within a binary lattice is denoted by v, the number of distribution patterns by r, and the number of DEGs by e. The degree of degeneration resulting from symmetry is shown to the right of each lattice. Binary lattices featuring the same number of case-containing cells are presented in the same row, with the total number of distribution patterns indicated at the end of each row.

As shown in the figure, when the subspace dimensionality equals one (d = 1), then the number of cells equals two (v = 2), and there is no reduction due to symmetry since there are no variables that can be exchanged; therefore, the number of DEGs is four (e = 4). When the subspace dimensionality equals two (d = 2), then the number of cells is four (v = 4), and the degree of reduction is two for the binary lattice in the middle of the second row due to the degeneration of different patterns by variable exchange. In this example, the total number of DEGs is twelve (e = 12).

4.3 Representation diversity of DEGs as relationships among variables

As stated in Subsection 2.2, to implement an FC generation function on a computer, pairs of subspaces sharing equivalent relationships must be chosen from high-dimensional data. Assuming this function is performed in the HCF, the number of d-dimensional DEGs (e) used to categorize relationships must exhibit diversity on the same order as the total number of subspaces ($S_N(d)$).

The excitatory neurons in the major subregions of HCFs consist of hundreds of thousands (dentate gyrus: about 800,000; CA1: about 400,000; CA3: about 300,000). Therefore, if we assume that variable N is the number of such cells, N falls within the

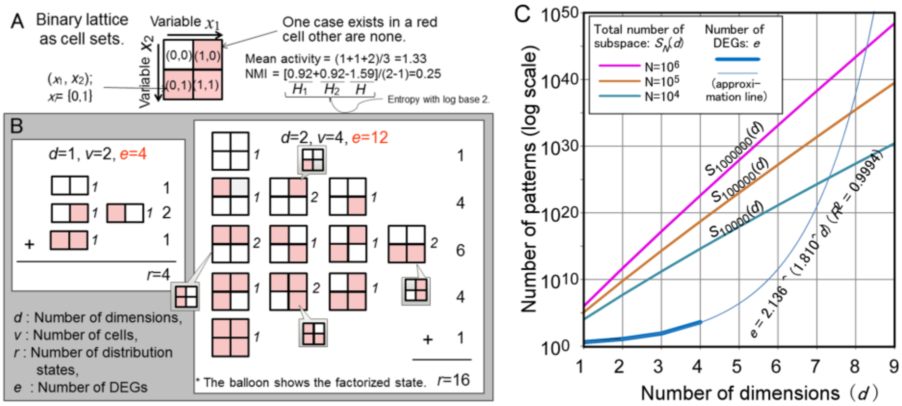

Fig. 2. Pattern of DEGs and number; A: Example of a binary lattice as a cell set, B: Pattern list of low dimensional DEGs, C: Number of DEGs corresponding to subspace dimensionality

range of 10^5 to 10^6. If using the 2-dimensional DEGs shown in Fig. 2-B as representations, their variety in pattern is only 12. Such a situation obviously lacks sufficient diversity to categorize the total number of subspaces, and as a result equivalent DEGs appear too frequently in many subspaces.

I examined the number of high dimensional DEGs (e) as shown in Fig. 2-C. Although the number of DEGs can actually be counted for DEGs with as many as four dimensions, it is not easy to do for DEGs with higher dimensionality. Thus, the number of DEGs with five or more dimensional subspaces were extrapolated by an approximation function, a second order exponential function of e = 2.136^(1.810^d). From this estimation, it was determined that 8-dimensional DEGs have a sufficient number of patterns to categorize the total number of subspaces extractable from more than 100,000 variables ($S_{100000}(d)$). In short, DEGs exhibit sufficient diversity to categorize all combination of subspaces within HCF; they have sufficient capacity for categorizing the relationships among variables in the system.

4.4 Appropriate DEGs within 4-dimensional patterns (4d-DEGs)

I next analyzed the appropriateness (in terms of predictability and mean activity level[5]) of the DEGs. In this section, I focus on the 3,984 patterns of 4-dimensional DEGs, which are highest dimensionality within those actually counted. Because of the length limit, I will omit an explanation of the algorithm for enumerating 4d-DEGs.

As mentioned in Subsection 2.2, the higher the predictability of newly generated FCs, the greater the chance of their being chosen in the following verification process. Generally, the higher the predictability of DEGs that connect variables in generated FCs, the higher the predictability of the generated FCs. Normalized mutual information (NMI) is estimated for every pattern of DEGs. Here, the NMI value is calculated by taking log base 2 and dividing by d-1 (Fig. 2-A); each cell is assumed to

[5] Activity level refers to the mean number of variables whose values equal 1. (Fig. 2-A)

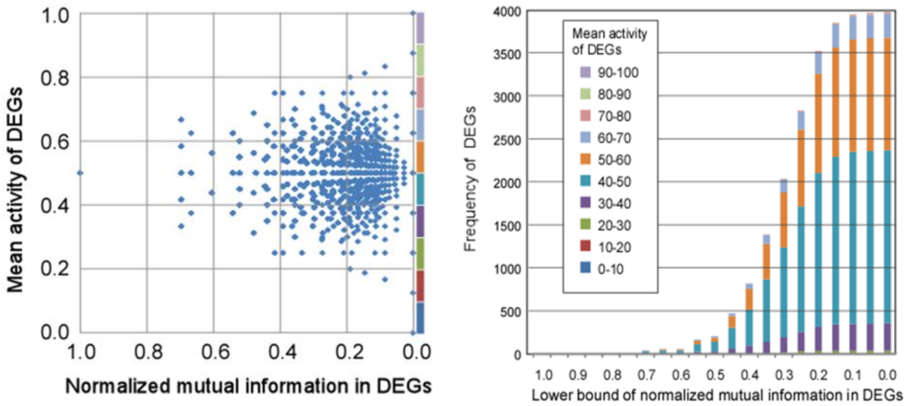

Fig. 3. Profiles of 4d- DEGs; A: Normalized mutual information (NMI) and mean activity of each DEG, B: Number of DEGs over the specified NMI value depending on the mean activity.

contain either one or zero cases. Because the information encoding of HCFs is thought to be sparse, I also calculated the mean activity[6] of each pattern.

NMI has a peak value of 1, which is obtainable only when the mean activity level is 0.5; the value decreases to 0 toward both ends as shown in Fig. 3-A. Fig. 3-B shows the frequencies of NMI that exceed certain values. From these figures, one can see that DEGs are densely distributed around a mean activity level of 0.5. There were only 203 DEG patterns with a highly predictable NMI of 0.5 or more. There were no DEGs whose NMI value was 0.35 or more and mean activity was 0.3 or lower.

If similar natures can be assumed for DEGs of higher dimensionality, we can conjecture regarding the next discussion. Although neural activity is sparse overall, these results suggest that relationships of variables are likely to be represented by locally active neuron groups such as cell assemblies[7].

5 Conclusions

Though the autonomous generation of domain-specific frame representations seems to be an indispensable function for realizing an AGI system, no technology exists to implement such a function. In designing a frame generation function, a subprocedure that is difficult to implement is the enumeration of frame candidates from the huge hypothesis space at a practical calculation cost. Assuming these candidate frames are generated by the cases-join process using variable assimilation, an index of representations that can express various relationships among variables will play an important role in this subprocedure. Sufficient scientific evidence exists to show that

[6] Mean activity refers to the mean of activity across all cases in one DEG pattern. Here, case activity refers to the rate of variables whose value is 1 for a single arbitrary case.

[7] Cell assembly is a diffuse structure comprising cells in the cortex and diencephalon; it is capable of acting briefly as a closed system and delivering facilitation to other such systems [17].

the neocortex is the region responsible for accumulating and activating frames. HCF is thought to be the region responsible for generating frames before they are transported to the neocortex. I conjectured that the information representation unique to HCF corresponds to the representations for relationships among variables. These representations are DEGs, each of which is a set of about ten cases in a multi-dimensional subspace that considers variable exchange symmetry. For simplicity, I used binary variable DEGs for estimating the diversity of representation, and my results showed such an approach exhibits sufficient diversity to categorize all combinations of subspaces within HCF. It has been suggested that DEGs are represented by active neuron groups such as cell assemblies.

Consequently, DEGs are a plausible candidate representation of the relationship among variables for a frame generation function; its algorithm can be studied based on this hypothesis in the future.

References

1. Minsky, M.: The Society of Mind, The Society of Mind (1988)
2. Fisher, D.H.: Knowledge Acquisition via Incremental Conceptual Clustering, Machine Learning, vol.2, pp.139-172 (1987)
3. Yamakawa, H., Maruhashi, K., and Nakao, Y.: Multi-Aspect Gene Relation Analysis, Pacific Symposium on Biocomputing 10:233-244(2005)
4. Newcombe, N., Picture This: Increasing Math and Science Learning by Improving Spatial, American Educator, Summer, pp.29-43 (2010)
5. Gentner, D.: Structure-mapping: Theoretical framework for analogy, Cognitive Science, Vol. 7, No. 2, pp.155-170 (1983)
6. Doumas. L. A. A., Hummel, J. E., & Sandhofer, C. M.: A theory of the discovery and predication of relational concepts. Psychological Review, 115, 1-43 (2008)
7. Wan, X., et. al.., The Neural Basis of Intuitive Best Next-Move Generation in Board Game Experts, Science, 331,6015, 341-346, (2011).
8. Yamakawa, H., What is neural basis of flexible inductive inferences?, Proc. JNNS, (2011)
9. Wagatsuma, N., et. al. , Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit model, Frontiers in Computational Neuroscience, 5, (2011)
10. Penner, M. R., Mizumori, S. J., Neural systems analysis of decision making during goal-directed navigation, Prog Neurobiol., 96(1): 96-135, 2012.
11. Bird, C. M., Burgess, N., The hippocampus and memory: insights from spatial processing, Nature Reviews Neuroscience 9, 182-194 (2008)
12. Johansson, C., Lansner, A., Towards cortex sized artificial neural systems, Neural Networks 20, 48-61 (2007)
13. Lisman, J. E., Role of the dual entorhinal inputs to hippocampus: a hypothesis based on cue/action (non-self/self) couplets, Prog Brain Res. 163:615-818 (2007)
14. Cohen, N. J., Eichenbaum, H., Memory, Amnesia, and the Hippocampal System, (1993)
15. Yamaguchi Y., A theory of hippocampal memory based on theta phase precession, Biol Cybern., 89(1), 1-9, (2003)
16. Rudy, J. W., & Sutherland, R. J. Configural and elemental associations and the memory coherence problem. J. Cognitive Neuroscience, 4, 208-216 (1992)
17. Hebb, D. O., The organization of behavior: a neuropsychological theory, John Wiley & Sons (1949).