Optimizing Potential Information Transfer with Self-referential Memory

Mikhail Prokopenko[1], Daniel Polani[2] and Peter Wang[1]

[1]CSIRO Information and Communication Technology Centre
Locked bag 17, North Ryde, NSW 1670, Australia
[2] Department of Computer Science, University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
mikhail.prokopenko@csiro.au

Abstract. This paper investigates an information-theoretic design principle, intended to support an evolution of a memory structure fitting a specific selection pressure: potential information transfer through the structure. The proposed criteria measure how much does associativity in memory add to the information transfer in terms of precision, recall and effectiveness. Maximization of the latter results in holographic memory structures that can be interpreted in self-referential terms. The study introduces an analogy between self-replication and memory retrieval, with DNA as a partially-associative memory containing relevant information. DNA decoding by a complicated protein machinery ("cues" or "keys") may corresponds to an associative recall: i.e., a replicated offspring is an associatively-recalled prototype. The proposed information-theoretic criteria intend to formalize the notion of information transfer involved in self-replication, and enable bio-inspired design of more effective memory structures.

1   Introduction

Bio-inspired models have been suggested and used in many areas of Unconventional Computing: parallel processing such as Cellular Automata (CA) and DNA computation; distributed storage and transmission: e.g., neural networks and associative memory; search and optimization: e.g., genetic algorithms and ant colony optimization (ACO). New metaphors are discovered and applied at an increasing pace, improving computational models in terms of robustness, adaptivity and scalability. However, there is a certain lack of a unifying methodology, or at least a set of guiding principles, underlying many recent developments. This is unsatisfactory not only from a methodological, but also from a pragmatic point of view: if some generic principles are not utilized then specific solutions are likely to be suboptimal.

Existence of such core principles may be supported by an observation that most of the bio-inspired models listed above do not fit into a particular category of conventional computing (memory, communication, processing), but cope with multiple aspects. For instance, CA were shown by Langton [18] to support,

under certain conditions (*the edge of chaos*), three basic operations of information storage, transmission, and modification, through static, propagating and interacting structures (*blinkers*, *gliders*, *collisions*). ACO algorithms also combine distributed memory, distributed transmission and distributed search, employing stigmergy — the process by which multiple ant-like agents indirectly interact through changes in their environment caused by pheromone deposits [4, 5] — and resulting in emergence of optimal solutions. In other words, these fundamental aspects of dealing with information are fused together within these bio-inspired approaches, making them less brittle and more scalable than conventional systems. One compelling explanation is that the motivating biological systems (ranging from cellular tissues to ant colonies) co-evolved the computing components rather than assemble the overall architecture out of separately designed parts [11].

The main question then becomes what are the core principles that interrelate memory, communication, and processing in evolvable computational systems? Answering this question from an information-theoretic viewpoint may also improve comparability of different bio-inspired approaches. In this paper, we propose an information-theoretic design principle, intended to support an evolution of a memory structure fitting a specific selection pressure: potential information transfer through the structure. In doing so we minimize architectural assumptions about memory or processor structures, hoping instead that such dependencies emerge as a result of the optimization of the information processing dynamics. Our preliminary studies, reported here, indicate that the proposed principle is capable of clearly identifying the range and information dynamics of possible memory structures in a very general sense, enabling design of optimal memory.

The following Section points out some relevant background material on unconventional memory organization, as well as intrinsic information-theoretic fitness criteria used in evolvable computational systems. Section 3 describes the proposed measure, followed by experimental results (Section 4) and conclusions (Section 5).

2 Background and Motivation

Moskowitz and Jousselin [20] have shown that, in a general algebraic sense, the nature of the operations carried out by a computer processor actually determine the structure of the computer memory. In particular, they highlighted the hidden group structure of the address space, and pointed out that "when the integer addition law is used to manipulate addresses, this space is a cyclic group, and memory is seen as a linear array". When another composition law is used (e.g., a non-commutative address composition), a hypercubic memory structure fits more, greatly reducing complexity of computations.

Another related concept is associative or content-addressed memory: a memory organization in which the memory is accessed by its content rather than an explicit address. Reference clues or keys are "associated" with actual mem-

ory contents until a desirable match (or set of matches) is found. A well-known example is a self-organizing map (SOM or Kohonen network). It can be interpreted as an associative memory which encodes the input patterns within the nodes of the network (the neural layer), in the form of weight (codebook) vectors of the same dimension and nature as the input patterns [17]. When a partial or corrupted pattern of data (a sensory cue) is presented in the form of a key input-vector, the rest of the pattern (memory) is associated with it. A characteristic of SOM-based associative memory is its self-organizing ordering: neighboring nodes encode similar codebook vectors, preserving topology: neurons that are closer in the neural layer tend to respond to inputs that are closer in the input space. A related approach is advocated by Kanerva [15, 16]: a Sparse Distributed Memory (SDM) which is a content addressable, associative memory technique relying on close memory items clustered together: while perceived data sparsely distribute themselves over multiple storage locations, the outcome is a fusion of this distribution. In the auto-associative version of SDM the memory contents and their addresses are from the same space and may be used alternatively. Another well-known example of auto-associative memory reproducing its input pattern as output is the Hopfield neural network [10].

Importance of memory access is discussed by Goertzel [6], who pursues "not a model of how memories are physically stored in the brain or anywhere else, but rather a model of how memory access must work, of how the time required to access different memories in different situations must vary". This pursuit led towards a *structurally associative memory* (STRAM), based on the idea that "if x is more easily accessible than y, those things which are similar to x should in general be more easily accessible than those things which are similar to y" [6]. Goertzel sketched a way of mapping a weighted graph describing STRAM to a physical memory M, by assigning to each pair of elements (x, y) stored by M a distance $D_M(x, y)$ measuring the difficulty of locating x in memory given that y has very recently been located. It was suggested that the distance $D_M(x, y)$ is approximated as a number of links along the shortest path between the graph nodes corresponding to x and y.

It is worth pointing out that our approach does not intend to present just a new measure of associativity or information transfer involved in memory operations, but rather identify an information-theoretic principle contributing to a general methodology. Such a methodology may go beyond computational aspects, including sensing, actuation, and networking in distributed systems, co-evolving under multiple design/selection pressures. We intend to follow the proposal on *information-driven evolutionary design* which suggested to use fitness functions according to generic information-theoretic criteria [25, 26, 13, 14]. The identification of possible intrinsic fitness criteria is also related to the work of Der *et al.* on self-organization of agent behaviors from domain-invariant principles, e.g., homeokinesis [3].

An example of a selection pressure is the acquisition of information from the environment: there is some evidence that pushing the information flow to the information-theoretic limit (i.e., maximization of information transfer in

perception-action loops) can give rise to intricate behaviour, induce a necessary structure in the system, and ultimately be responsible for adaptively reshaping the system [12–14]. Other important selection pressures applicable to distributed systems include stability of self-organizing hierarchies [22]; efficiency of multi-cellular communication topologies [23]; efficiency of locomotion and distributed actuation [25, 29]. Identifying a selection pressure on potential information transfer involved in memory recall would allow us to contribute to the general methodology of information-driven evolutionary design.

3 Information Transfer: Precision and Recall

Since our task is to identify a very generic principle, we choose to abstract away from implementation details and consider instead an unconstrained deterministic function f from two equally distributed random variables K and X to a random variable Y. The variable K is intended to serve as a "key" or "cue" in accessing the memory X, retrieving, as a result of the mapping f, the outcome Y, i.e., $Y = f(K, X)$. It is important to realize that while we interpret K, X and Y as key, memory and outcome, we do not structurally constrain the variables and the mapping: e.g., there is no requirement that any location x in memory X is accessible by a unique key $k \in K$, etc.

The first constraint that we impose is the criterion:

$$\text{maximization of} \quad \mathcal{P} = I(X; K|Y) , \tag{1}$$

where $I(A; B)$ denotes the mutual information between A and B:

$$I(A; B) = \sum_{a \in A} \sum_{b \in B} P(a, b) \, \log \frac{P(a, b)}{P(a)P(b)} , \tag{2}$$

where $P(a)$ is the probability that A is in the state a, and $P(a, b)$ is the joint probability. The criterion (1) maximizes the conditional mutual information between key and memory, given the outcome. First of all, we need to clarify that, although K and X are independent and, therefore, mutual information $I(X; K)$ is zero, the conditional mutual information $I(X; K|Y)$ may well be positive. This is analogous to the example of a binary symmetric channel with input X, noise K, and output Y, described by MacKay [19] (we altered the variables names here to avoid confusion): mutual information $I(X; K) = 0$ since input and noise are independent, but $I(X; K|Y) > 0$, because "once you see the output, the unknown input and the unknown noise are intimately related!" [19]. Similarly, the criterion (1) is applied once the outcome is obtained, which means that a possible association between memory and key has been made.

Secondly, we draw an analogy with well-known information retrieval metrics: precision and recall. Precision is a measure of usefulness or *soundness* of the outcome retrieved in response to a query, and is measured as a fraction of the relevant and retrieved items within the retrieved items (aiming at "nothing but the truth"). Recall is a measure of relevance or *completeness* of the retrieved

outcome, and is measured as a ratio of the relevant and retrieved items over the relevant items (aiming at "the whole truth"). A probabilistic interpretation is possible as well [7]: precision may be defined as the conditional probability that an object is relevant given that it is returned by the system, while the recall is the conditional probability that a relevant object is returned: precision = $P(\text{relevant}|\text{returned})$, and recall = $P(\text{returned}|\text{relevant})$.

Intuitively, the criterion (1) captures the potential $\mathcal{P}$ of precision-driven information transfer. To formalize this intuition, let us apply the chain rule for the mutual information:

$$I(X; Y, K) = I(X; Y) + I(X; K|Y) ,$$

producing

$$\mathcal{P} = I(X; K|Y) = I(X; Y, K) - I(X; Y) . \tag{3}$$

The alternative representation (3) can be interpreted as follows: how much does a key *add to precision of the outcome by associating with memory*. The equation (3) contrasts two information transfers: one, $I(X; Y)$, does not use associativity, while the other, $I(X; Y, K)$, incorporates it. The difference between the two transfers captures, we believe, the potential information gain in precision. Another useful representation of the criterion (1) can be obtained in terms of entropies $H(\cdot)$, joint entropies $H(\cdot, \cdot)$, and conditional entropies $H(\cdot|\cdot)$:

$$H(A) = - \sum_{a \in A} P(a) \, \log P(a) , \tag{4}$$

$$H(A, B) = - \sum_{a \in A} \sum_{b \in B} P(a, b) \, \log P(a, b) , \tag{5}$$

$$H(B|A) = H(A, B) - H(A) , \tag{6}$$

where $P(a)$ is the probability that A is in the state a, $P(b)$ is the probability that B is in the state b, and $P(a, b)$ is the joint probability. We begin by applying the identity

$$I(X; K|Y) = H(X|Y) - H(X|K, Y)$$

to the right-hand side of the criterion (1). It yields

$$\mathcal{P} = I(X; K|Y) = H(X|Y) - H(X|K, Y) =$$

$$H(X|Y) - [H(X, Y, K) - H(K) - H(Y|K)] =$$

$$H(X|Y) - H(X, Y, K) + H(K) + [H(K, Y) - H(K)] =$$

$$[H(X, Y) - H(Y)] - H(X, Y, K) + H(K, Y) .$$

where the last three steps used relationships $H(X|K, Y) = H(X, Y, K) - H(K) - H(Y|K)$, $H(Y|K) = H(K, Y) - H(K)$ and $H(X|Y) = H(X, Y) - H(Y)$ respectively. A further reduction is possible for deterministic functions, where $H(X, Y, K)$ is a constant, making the criterion (1) equivalent to

$$\text{maximization of} \quad \tilde{\mathcal{P}} = H(X, Y) - H(Y) + H(K, Y) . \tag{7}$$

The measure $\tilde{\mathcal{P}}$ may, of course, be rewritten as follows:

$$\tilde{\mathcal{P}} = H(X|Y) + H(K,Y) = H(X,Y) + H(K|Y) . \tag{8}$$

At this stage we would like to introduce another criterion. We consider

$$\text{maximization of} \quad \mathcal{R} = I(Y;K|X) = I(Y;X,K) - I(Y,X) . \tag{9}$$

Intuitively, $\mathcal{R}$ measures how much a key is necessary *to identify the output of the mapping, given the memory*. The criterion (9) captures the potential $\mathcal{R}$ of information transfer involved in the memory recall, and aims to maximize the difference between associative and non-associative information transfer. Using similar unfolding, we obtain

$$\mathcal{R} = I(Y;K|X) = H(Y|X) - H(Y|K,X) =$$

$$H(Y|X) - [H(Y,X,K) - H(K) - H(X|K)] =$$

$$H(Y|X) - H(Y,X,K) + H(K) + [H(X,K) - H(K)] =$$

$$[H(Y,X) - H(X)] - H(Y,X,K) + H(X,K) .$$

Since for deterministic functions the last three entropies of the right-hand side are constants (and $H(Y,X,K) = H(X,K)$ anyway), maximization of $\mathcal{R}$ is equivalent to

$$\text{maximization of} \quad \tilde{\mathcal{R}} = H(X,Y) . \tag{10}$$

It should be noted that since $Y = f(K,X)$, the expression for $\tilde{R}$ is dependent on K.

The overall effectiveness of information retrieval is typically defined as the harmonic mean (the reciprocal of the arithmetic mean of the reciprocals) of recall and precision — hence, we suggest the criterion:

$$\text{maximization of} \quad \mathcal{E} = \frac{2}{\frac{1}{\mathcal{P}} + \frac{1}{\mathcal{R}}} = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}} , \tag{11}$$

fusing together the potential information gains in both precision and recall.

In order to highlight different roles played by K and X, we consider here scenarios with varying sizes $\|K\|$ and $\|X\|$, interpreted in the context of several examples: (a) catalog/book indexing and search; (b) pattern association using a neural network; (c) decoding of genotype (DNA) by proteins. The scenarios are as follows:

(S_1) $\quad \|K\| \gg \|X\|$ and $\|X\| \approx \|Y\|$, $\qquad (S_2)$ $\quad \|K\| \approx \|X\| \approx \|Y\|$,
(S_3) $\quad \|K\| \ll \|X\|$ and $\|X\| \approx \|Y\|$

where $\| \circ \|$ is the cardinal number of the set $\circ$ (in our case, simply the number of its elements). In the example (a), a library catalogue is a database containing records indexed by the authors, titles, subjects, etc. The explicit "cue" is the key, using which a set of catalogue items $Y' \subseteq Y$ can be found as a result of a query. Typically, $\|K\| \gg \|X\|$, while $\|Y\| \approx \|X\|$: this is our first scenario (S_1).

Similarly, a book can be indexed by associating its content (e.g., pages) with keywords. In this case, $\|K\| \gg \|X\|$ as well, since there may be more keywords than pages, while $\|Y\| \approx \|X\|$ as the number of retrieved pages may approach their total number. However, the scenario (S_2) pushes the scenario (S_1) to the extreme by restricting the number of possible keys (e.g., a limit on queries), while the memory size is unchanged: $\|K\| \approx \|X\|$. This represents a more challenging case with respect to the precision as the relevant items are harder to find.

The example (b) involves an artificial neural network, e.g., a self-organizing map (SOM) implementing associative memory, briefly discussed in section 2. Each neuron in memory (a network node) encodes a retrievable pattern, hence $\|Y\| \approx \|X\|$. Of course, memory updates would lead to an increase in the overall number of returned patterns, highlighting the distinction between cumulative memory capacity and memory size. The SOM handles multiple cues/keys as partial or corrupted patterns of data, associating them with the memory, implying $\|K\| \gg \|X\|$. This also concurs with the first scenario (S_1). Again, restricting the number of possible keys while keeping the memory size is unchanged (the scenario (S_2)) would challenge the system in terms of the precision.

The third scenario (S_3) may correspond to an auto-associative neural network such as the Hopfield network [10] or a Sparse Distributed Memory [15]. A key is interpreted simultaneously by all neurons which interact by updating their weights until a stable network state is reached: this attractor then represents the network output associated with the key. In this case, $\|X\| \gg \|Y\|$ since there is only a limited number of attractor states supported by the network, while $\|K\| \ll \|X\|$ due to high-dimensionality of memory. Interestingly, restricting the memory (reducing $\|X\|$) would challenge precision again, approaching the scenario (S_2) from another direction.

Finally, we consider the case (c) when a genotype (DNA) is decoded by proteins. An individual DNA can be interpreted as associative memory in the sense that it contains *potential information* relevant to the niche occupied by the individual's species. As pointed out by Adami [1], "If you do not know which system your sequence refers to, then whatever is on it cannot be considered information. Instead, it is potential information (a.k.a. entropy)". Decoding a DNA involves a complicated protein machinery (the key), and may corresponds to an associative recall. In this model, a replicated offspring is an associatively-recalled prototype. In the next section we shall interpret all three scenarios within this analogy.

4 Results

The experimental setup is very simple: we intend to satisfy our criteria (1), (9), and (11) by varying possible deterministic functions $Y = f(K, X)$ over finite size domains K, X and Y, for the scenarios (S_1), (S_2) and (S_3). In particular, we consider three sets of integers $\{1, \ldots, \|K\|\}$, $\{1, \ldots, \|X\|\}$ and $\{1, \ldots, \|Y\|\}$, and vary their sizes $\|K\|$, $\|X\|$ and $\|Y\|$ between experiments. For each experiment, we search for deterministic mappings $Y = f(K, X)$ which maximize $\mathcal{P}$,

or $\mathcal{R}$, or $\mathcal{E}$ — repeating the search for each of these criteria. We used a simple genetic algorithm to evolve solutions to the maximization problems. The initial population is generated by random mappings $Y = f_i(K, X)$, for a sufficiently large number of individual mappings, e.g. $1 \leq i \leq 1000$. At each generation, the mappings are evaluated in terms of the criterion in point (either $\mathcal{P}$, or $\mathcal{R}$, or $\mathcal{E}$). We have chosen a generation gap replacement strategy (the entire old population is sorted according to the fitness, and the best 10% are chosen for direct replication in the next generation, employing an elitist selection mechanism), and the multiple-point crossover. We also ensure that the mutation results in a unique individual by re-applying this operator if necessary.

Visualizing evolved mappings f is not revealing, as can be observed from Figure 1. We plot instead an analogue of a 2-dimensional contour, but rather than simply using contours, we connect, for a given height $y \in Y$, all points $(k, x) \in K \times X$ which agree either on k or on x, producing a partial grid. For example, if there are entries $7 = f(1, 4)$, $7 = f(3, 4)$, and $7 = f(1, 6)$, we connect points $(1, 4)$ and $(3, 4)$ as they represent the same memory $x = 4$, as well as points $(1, 4)$ and $(1, 6)$ sharing the same key $k = 1$. A grid-contour combines grids for all values of $y \in Y$ by "overlaying" the grids for all values y.

A random mapping (the zero hypothesis) has no discernable structure for all scenarios (e.g., Figure 2). Let us focus initially on the scenario (S_1). A $\mathcal{P}$-maximizing mapping for this scenario is a structure with dominant horizontal lines (Figure 3). Each horizontal reflects the fact that in the evolved mapping, the same memory is recalled if multiple different keys are associated with it. This, in the context of DNA decoding, corresponds to conservation of DNA (memory) and its robustness to possible decoding errors (multiple keys), ensuring high precision. A $\mathcal{R}$-maximizing mapping maintains the horizontal lines but introduces some vertical lines (Figure 4). Each vertical line means that a key recalls the same content even if associated with different memories. In the context of DNA decoding, this would correspond to junk DNA: redundant code which does not differentiate between offsprings and ensures high recall. Importantly, the effectiveness criterion $\mathcal{E}$ maintains the horizontal lines (robust DNA) but eliminates the vertical lines (no junk DNA), as shown in Figure 3. On the other hand, minimization of $\mathcal{E}$ does the opposite, producing a grid-like structure, i.e., for every association (k_1, x_1) there exists an association (k_2, x_2) such that either $k_1 = k_2$ or $x_1 = x_2$.

The scenario (S_2) pushes the observed tendencies to their limits. A $\mathcal{P}$-maximizing mapping for this scenario is a structure with no lines (Figure 5). There are no entries which share either a key or memory — in other words, both key and memory are necessary. Such a *holographic* outcome illustrates the full precision of associative memory (a perfectly succinct DNA). An $\mathcal{R}$-maximizing mapping has some vertical lines (Figure 6), suggesting that some junk DNA is possible even in the highest recall case. This can be interpreted as a tendency towards the dominance of precision over recall, i.e., robustness of DNA at the expense of redundancy. However, the effectiveness criterion $\mathcal{E}$ eliminates redundancy and results in the holographic structure (Figure 5). This mapping implements a fully

associative memory, where for every pair of a key and memory, fixing a key k and varying memory x (or vice versa) results in a different outcome $y = f(k, x)$.

The results for the scenario (S_3) are not surprising: mappings maximizing $\mathcal{P}$, $\mathcal{R}$ and $\mathcal{E}$ produce structures with only vertical lines. In the context of DNA decoding, this would correspond to highly redundant and error-prone DNA structures. This model would work for reproduction if different arrays collectively store information (as in an SDM or Hopfield network) "retrieving" offspring as a composite result of data fusion, e.g. genetic cross-over.

Fig. 1. An evolved mapping: scenario (S_2).

A holographic mapping $Y = f(K, X)$ implementing fully associative memory in the scenario (S_2) (Figure 5) can be interpreted in self-referential terms. Self-referentiality has many interpretations, ranging from programming data structures (a self-referential structure contains a pointer to a structure of the same type) to cognitive neuroscience: the self is a cognitive structure with special mnemonic abilities, leading to "the enhanced memorability of material processed in relation to self" [8, 27], suggesting that a self-referential memory — a memory about the self — is not ordinary. According to the well-known interpretation of Hofstadter [9], a self-referential system can be characterised by emergent behaviour and tangled hierarchies exhibiting Strange Loops: "an interaction between levels in which the top level reaches back down towards the bottom level
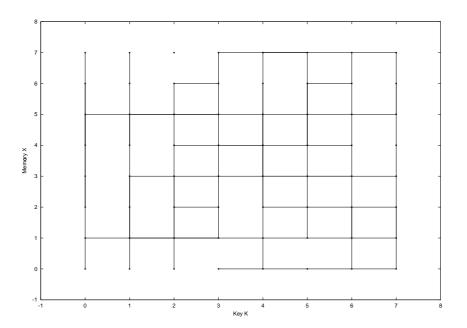
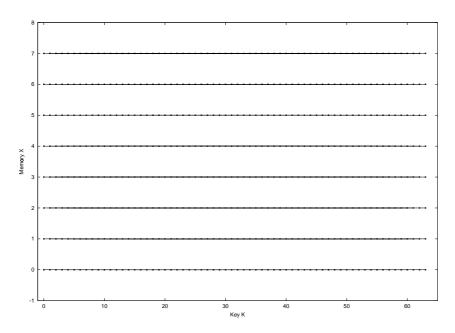**Fig. 2.** Grid-contour of a random mapping: scenario (S_2).

Fig. 3. Grid-contour of a $\mathcal{P}$-maximizing mapping, as well as an $\mathcal{E}$-maximizing mapping: scenario (S_1).
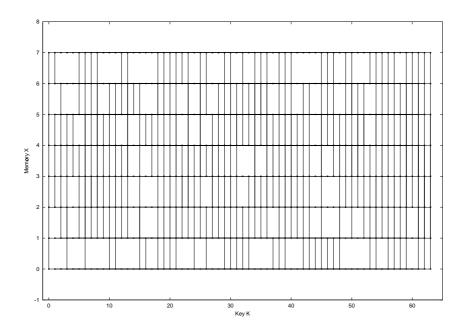
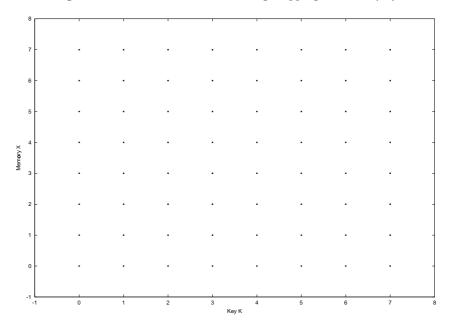**Fig. 4.** Grid-contour of a $\mathcal{R}$-maximizing mapping: scenario (S_1).

Fig. 5. Grid-contour of a holographic $\mathcal{P}$-maximizing mapping, as well as an $\mathcal{E}$-maximizing mapping: scenario (S_2). Its 3-dimensional counterpart is shown in Figure 1.
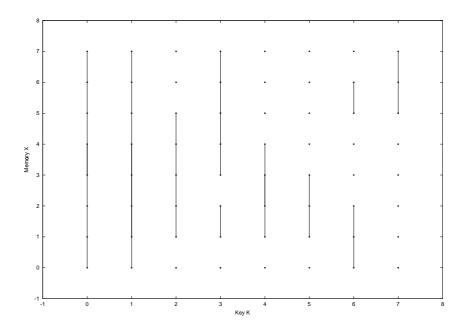
Fig. 6. Grid-contour of a $\mathcal{R}$-maximizing mapping: scenario (S_2).

and influences it, while at the same time being itself determined by the bottom level". We shall adopt a weaker interpretation of self-referential memory: the memory using a model of itself. This limited a-model-within-the-model view is not intended to preclude emergence of tangled hierarchies, or references to the cognitive self of the agent using this memory.

We begin by observing that, on the one hand, for $y_{1,1} = f(k_1, x_1)$ and $y_{1,2} = f(k_1, x_2)$ (i.e., the same key and varying memory), the outcomes differ: $y_{1,1} \neq y_{1,2}$. This means that every memory is sufficiently sensitive to its own content/location, and there is no redundant information in the associated key: the difference in the recall is due to different memory. On the other hand, for $y_{1,1} = f(k_1, x_1)$ and $y_{2,1} = f(k_2, x_1)$ (i.e., varying key and the same memory), the outcomes differ as well: $y_{1,1} \neq y_{2,1}$. This means that every key is sufficiently informative to produce different outcomes upon association with the same memory. In other words, every memory content is sufficiently sensitive to each key (as well as to its content/location), and therefore, encodes information about all possible keys. If the memory was simply mirroring each associated key, it would not be sensitive to its own content/location. Hence, each memory content *uniquely* encodes information about all possible keys, e.g., in an array of permuted keys. This information *per se* is a memory model. Hence, each memory content maintains a possible model of itself. We believe that it is precisely this self-referentiality that leads to *holographic-ness* and optimizes effectiveness of the considered information transfer in terms of precision and recall. The self-

referentiality emerges under the pressures imposed by restricting the number of queries and outcomes to the memory size: the scenario (S_2). If one of these pressures is relaxed, self-referentiality is not needed and a memory does not have to encode information about all possible keys: hence, the presence of horizontal lines in the optimal structures for the scenario (S_1), or vertical lines for the scenario (S_3).

5   Discussion and Conclusions

The evolved mappings $Y = f(K, X)$ maximize precision, recall and effectiveness of the potential information transfer throughout associative memory. Maximization results in holographic structures that can be interpreted in self-referential terms. On the contrary, minimization of the proposed criteria result in full-grid structures which do not require associativity as it does not affect the involved information transfer.

It was conjectured (e.g., [21, 24]) that the degree of self-referentiality employed by a self-replicating multi-cellular organism is related to efficiency of its self-inspection and self-repair — and may be quantitatively measured in order to evolve more efficient processes. This conjecture was tested in this work in terms of memory structures and the information transfer. Continuing our analogy with DNA as an associative memory, it is interesting to observe that real-life examples of DNA are not approaching the maximum information transfer, as evidenced by their non-perfect error recovery and significant redundancy. Thus, in terms of self-replication, the maximum potential is not realized — it would require higher precision and higher recall, culminating in a perfectly-associative memory. Interestingly, another extreme, lower precision and/or lower recall, can be pointed out already. We believe that a suitable example is the self-replication mechanism exhibited by mineral crystals in the absence of biological enzymes, as advocated by Cairns-Smith [2]: clay crystals can store information as a pattern of inhomogeneities that are propagated from layer to layer, with few errors; they can reproduce by random fragmentation; and they can express a variety of morphological phenotypes. Following this intuition, Schulman and Winfree recently proposed a method of error-correcting self-replication that works by similar growth and fragmentation of algorithmic DNA crystals [28]: "crystal growth extends the layers and copies the sequence of orientations, which may be considered its genotype. ... splitting of a crystal can yield multiple pieces, each containing at least one copy of the entire genotype". Such self-replication can be considered as non-associative memory recall, where a key is not necessary at all, and neither the point of crystal fragmentation nor surrounding environmental conditions are important. In other words, Cairns-Smith model of crystal self-replication is near the low-precision and low-recall extreme, while a self-referential associative memory would implement the highest-effectiveness case.

Adami advocated the view that "evolution increases the amount of information a population harbors about its niche" [1]. The information-theoretic criteria

proposed in this work may further formalize the notion of information transfer involved in self-replication, and enable bio-inspired design of more effective memory structures.

References

1. Adami, C., 2002, What is complexity?, *Bioessays*, 24(12), 1085–1094.
2. Cairns-Smith, A.G., 1966, The origin of life and the nature of the primitive gene, *Journal of Theoretical Biology*, 10, 53–88.
3. Der, R., Steinmetz, U., Pasemann, F., 1999, Homeokinesis — A new principle to back up evolution with learning, *Concurrent Systems Engineering Series*, 55, 43–47.
4. Dorigo, M., Maniezzo, V., Colorni, A., 1996, The Ant System: Optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man, and Cybernetics, B*, 26(1), 1–13.
5. Dorigo, M., and Di Caro, G., 1999, Ant Colony Optimization: A new meta-heuristic, in: *Proceedings of 1999 Congress on Evolutionary Computation, Washington DC*, 1470–1477, IEEE Press, Piscataway, NJ.
6. Goertzel, B., 1993, *The structure of intelligence: A new mathematical model of mind*, New York: Springer-Verlag.
7. Goutte, C., and Gaussier, E., 2005, A probabilistic interpretation of precision, recall and F-score, with implication for evaluation, in: *Proceedings of the 27th European Conference on Information Retrieval*, Santiago de Compostela, Spain.
8. Heatherton, T.F., Macrae, C.N., Kelley, W.M., 2004, What the social brain sciences can tell us about the self, *Current Directions in Psychological Science*, 13(5), 190–193.
9. Hofstadter, D. R., 1989, *Gödel, Escher, Bach: An eternal golden braid*, New York: Vintage Books.
10. Hopfield, J.J., 1982, Neural networks and physical systems with emergent collective computation abilities, *Proceedings of the National Academy of Science*, 79, 2554–2558.
11. Goldsmith, R.S., and Miller, J.F., 2003, Cooperative co-evolution of robot control, sensor relevance and placement, in: *Proccedings of EPSRC Evolvability and Sensor Evolution Symposium*, Birmingham, U.K.
12. Klyubin, A. S., Polani, D., Nehaniv, C. L., 2004, Organization of the information flow in the perception-action loop of evolved agents, in: *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, IEEE Computer Society, 177–180.
13. Klyubin, A. S., Polani, D., and Nehaniv, C. L., 2005. Empowerment: A universal agent-centric measure of control, in: *Proceedings of The 2005 IEEE Congress on Evolutionary Computation*, IEEE Press, 128–135.
14. Klyubin, A. S., Polani, D., Nehaniv, C. L., 2005, All else being equal be empowered, in: Capcarrère, M.S., Freitas, A.A., Bentley, P.J., Johnson, C.G., Timmis, J. (eds.), *Advances in Artificial Life, 8th European Conference, ECAL, 2005, Canterbury, UK, September 5-9, 2005, Proceedings*, 744–753, Springer, Lecture Notes in Computer Science, Vol. 3630.
15. Kanerva, P., 1988, *Sparse Distributed Memory*, Cambridge, Mass.: MIT Press.
16. Kanerva, P., 1993, Sparse Distributed Memory and related models. In M.H. Hassoun (ed.), *Associative Neural Memories: Theory and Implementation*, New York: Oxford University Press, 50–76.

17. Kohonen, T., 1984, *Self-organization and associative memory*, Berlin, Springer-Verlag.

18. Langton, C., 1991, Computation at the Edge of Chaos: Phase transitions and emergent computation. In S. Forest (ed.), *Emergent Computation*, MIT.

19. MacKay, D. J. C., 2003, *Information theory, inference and learning algorithms*, Cambridge University Press.

20. Moskowitz, J.P., and Jousselin, C., 1989, An algebraic memory model. *ACM SIGARCH Computer Architecture News archive*, 17(1), 55–62, ACM Press, New York, NY, USA.

21. Prokopenko, M., and Wang, P., 2004, On Self-referential shape replication in robust aerospace vehicles, in Pollack, J. (ed.) *Artificial Life IX: Proceedings of The 9th International Conference on the Simulation and Synthesis of Living Systems*, Boston, USA, MIT Press, 27–32.

22. Prokopenko, M., Wang, P., Price, D.C., Valencia, P., Foreman, M., Farmer, A.J., 2005, Self-organizing hierarchies in sensor and communication networks. *Artificial Life*, Special Issue on Dynamic Hierarchies, 11(4), 407–426.

23. Prokopenko, M., Wang, P., Foreman, M., Valencia, P., Price, D.C., Poulton, G. T., 2005, On connectivity of reconfigurable impact networks in Ageless Aerospace Vehicles. *The Journal of Robotics and Autonomous Systems*, 53(1), 36–58.

24. Prokopenko, M., Poulton, G., Price, D. C., Wang, P., Valencia, P., Hoschke, N., Farmer, A. J., Hedley, M., Lewis, C., Scott, D. A., 2006, Self-organising impact sensing networks in robust aerospace vehicles, in Fulcher, J. (ed.) *Advances in Applied Artificial Intelligence*, Idea Group, 186–233.

25. Prokopenko, M., Gerasimov, V., Tanev, I., 2006, Measuring spatiotemporal coordination in a modular robotic system, in: Rocha, L.M., Yaeger, L.S., Bedau, M.A., Floreano, D., Goldstone, R.L., Vespignani, A., (eds.). *Artificial Life X: Proceedings of The 10th International Conference on the Simulation and Synthesis of Living Systems*, Bloomington IN, USA.

26. Prokopenko, M., Gerasimov, V., Tanev, I., 2006, Evolving spatiotemporal coordination in a modular robotic system, in *Proceedings of The 9th International Conference on the Simulation of Adaptive Behavior* (to appear), Rome, Italy.

27. Rogers, T.B., Kuiper, N.A., Kirker, W.S., 1977, Self-reference and the encoding of personal information, *Journal of Personality and Social Psychology*, 35(9), 677–688.

28. Schulman, R., and Winfree, E., 2005, Self-replication and evolution of DNA crystals, in: M. S. Capcarrère, A.A. Freitas, P.J. Bentley, C.G. Johnson, J, Timmis (eds.) *Advances in Artificial Life, 8th European Conference, ECAL 2005, Canterbury, UK, September 5-9, 2005, Proceedings*, 734–743, Springer.

29. Tanev, I., Ray, T., Buller, A., 2005, Automated evolutionary design, robustness, and adaptation of sidewinding locomotion of a simulated snake-like robot, *IEEE Transactions On Robotics*, 21, 632–645.