Self-Replicating Worms That Increase Structural Complexity through Gene Transmission

Hiroki Sayama
New England Complex Systems Institute
24 Mt. Auburn St., Cambridge, MA 02138
sayama@necsi.org

Abstract

A new self-replicating cellular automata (CA) model is proposed as a latest effort toward the realization of an artificial evolutionary system on CA where structural complexity of self-replicators can increase in some cases. I utilize the idea of 'shape encoding' proposed by Morita and Imai (Morita & Imai 1996b) and make the state-transition rules of the model allow organisms to transmit genetic information to others when colliding against each other. Simulations with random initial configuration demonstrate that it is possible that the average length of organisms and the average frequency of brancing per organism both increase, with decreasing self-replication fidelity, and saturate at some constant level. The saturation is caused in part by the fixation of place and shape of organisms onto particular sites. This implies the necessity of introducing some fluidity of site arrangements into the model for further development of evolutionary models using CA-like artificial media.

Introduction

Research on self-replicating patterns on CA was founded by von Neumann (von Neumann 1966) and now is viewed as one of the origins of artificial life research (Marchal 1998). A number of attempts to embody artificial organisms on CA have been conducted so far in this area. They may be categorized into four groups as follows[1]:

(1) Implementation of universal constructors based on von Neumann's self-reproducing automaton, studied in '50s–'70s (von Neumann 1966; Codd 1968; Vitányi 1973; Pesavento 1995)

(2) Search for a minimal system capable of non-trivial self-replication, studied in '80s–'90s (Langton 1984; Byl 1989; Reggia *et al.* 1993; Sipper 1994; Morita & Imai 1996b)

(3) Addition of other computational capabilities to self-replicators, studied in '90s–present (Tempesti 1995; Perrier, Sipper, & Zahnd 1996; Chou & Reggia 1998)

(4) Realization of emergence and evolution of self-replicators, studied in '90s–present (Lohn & Reggia 1995; 1997; Chou & Reggia 1997; Sayama 1998a; 1998b; 2000)

(1), (2), and (3) are efforts to implement regulated behavior (e.g. construction, self-replication, computation) manually designed according to the designer's idea, while (4) strives to obtain unexpected behavior (e.g. emergence of self-replicators or evolution) that may arise from robust or random state-transition rules.

In terms of category (4), the evolutionary processes so far attained using CA were just a change of the size of self-replicating loops, i.e. either increase (Chou & Reggia 1997) or decrease (Sayama 1998a; 2000) in size of the loops. Whether the complexity-increasing evolution of artificial organisms (the evolution of their ability to do more complicated things) is attainable using CA is an open question originally posed by von Neumann at the beginning of this area (von Neumann 1966; Marchal 1998), which still has been unsolved. One of the reasons for this is that the idea of Langton's self-replicating loop (Langton 1984) that has formed the basis for many succeeding studies requires a simple, square (or rectangular) shape of organisms to enable their replication. To remove this restriction, it is necessary to employ a model much more flexible in terms of the shape of self-replicating organisms.

The work introduced in this article extends this effort to realize a new CA model where complexity of virtual organisms can increase along time. I mainly focus on the possibility of increase in structural complexity of artificial organisms, based on the assumption that any other aspects of complexity such as the function of artificial organisms should stem from their structure. To construct a new CA model, I employ the

[1] See also (Sayama 1998a, Chap.3) and (Sipper 1998).

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

shape-encoding mechanism proposed by Morita and Imai (Morita & Imai 1996b) that makes a variety of patterns capable of self-replication. I then make the state-transition rules of the model allow organisms to transmit genetic information to others when colliding against each other, which may give rise to their variation.

In the following sections I introduce the design of the new model and demonstrate through simulations that the average length and the average frequency of brancing per organism can both increase in this model. Such processes take place with the decrease of self-replication fidelity due to overcrowding, and always saturate at some constant level. It is suggested that such saturation is caused in part by the fixation of place and shape of patterns onto particular sites, which is an intrinsic limitation of CA that prevents us from creating open-ended evolution there.

Model

The shape-encoding mechanism

The main property to be added to the CA-based self-replication model is the variety in shape of self-replicators. For this purpose, I utilize the shape-encoding mechanism proposed by Morita and Imai (Morita & Imai 1996b), which is a mechanism to let an organism dynamically generate genetic codes from its own phenotypical pattern by self-inspection (Laing 1977). An example of their self-replicating automata is shown in Fig. 1. This worm performs a unique form of self-replication, in which its shape is continuously being encoded into genotype at its tail, and these encoded genes are conveyed to the head and decoded there for construction of its offspring. This is different from other prevalent systems in which genetic information is described in a static form and information flows only from genotype to phenotype.

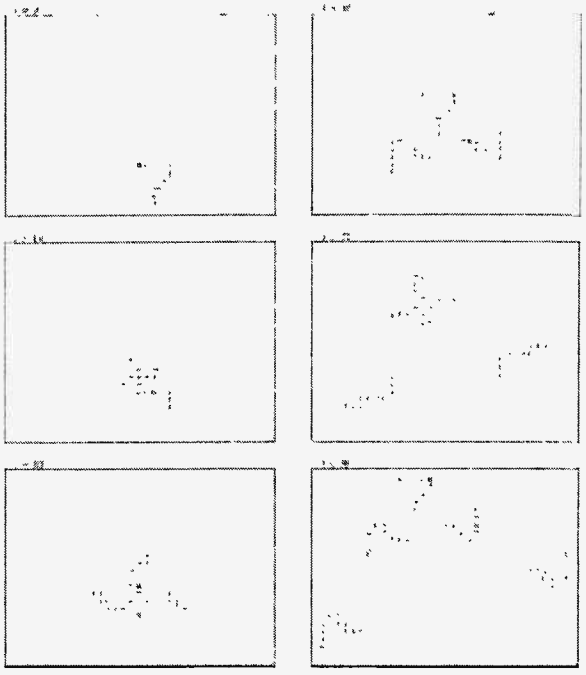

Figure 1: A self-replicating worm with the shape-encoding mechanism by Morita and Imai (from (Morita & Imai 1996b) by courtesy of the original authors).

With the shape-encoding mechanism, a great variety of patterns can self-replicate, which is quite useful for the goals of this study. Since Morita and Imai did not consider collisions among organisms, however, it is necessary to incorporate the state-transition rules for such situations.

Space

The proposed model uses a two-dimensional discrete space, where a virtual organism is represented as a contiguous region composed of mutually connected 'structure cells[2].' A structure cell is a square that has one input port and three output interfaces on its edge, and also three internal 1-bit registers inside itself, as shown in Fig. 2. The internal registers hold information about the existence of Central, Left, and Right genes that sequentially describe how the shape of that organism is formed. They are continuously being updated to have new values coming from the input port, while the old values they previously had are sent out through the output interfaces. The values of internal registers are conveyed successfully if and only if the cell's input port is correctly rooted to one of the output interfaces of another adjacent structure cell. In addition, each entire cell takes either active or passive mode. A passive cell simply conveys genetic information, while an active cell plays more important roles in growth and dissolution of the structure patterns.

The above-mentioned cell property is implemented using sixty-five-state CA with von Neumann neighborhood (five-site neighborhood), where the state each site will take after one update is determined locally according to the states it and its four adjacent sites (upper, lower, right, and left neighbors) have at present. The design of states used here is shown in Fig. 3. The states

[2]Note that the word 'cell' is used here for a particular state as a part of a virtual organism represented on the CA space, while 'site' is used for a substrate unit that composes the space itself.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

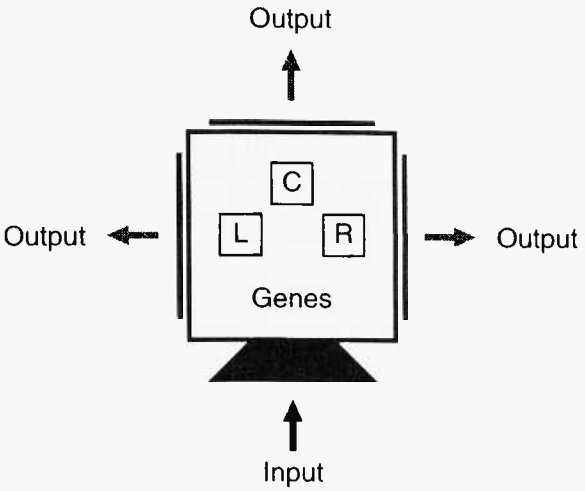

Figure 2: Schematic illustration of a structure cell in the new model. It has one input port, three output interfaces, and three internal 1-bit registers that hold information about the existence of Central, Left, and Right genes. The entire cell takes either active or passive mode.

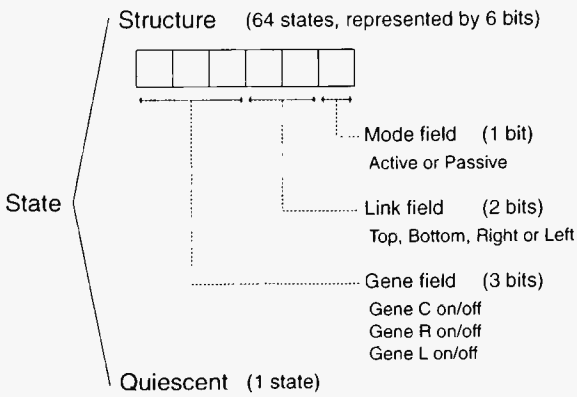

Figure 3: Design of states in the new model.

consist of one quiescent state and sixty-four structure states. The latter are composed of three parts: Mode field (1 bit), Link field (2 bits) and Gene field (3 bits). This implementation is based on the idea of 'multi-data-field CA' (Chou & Reggia 1997) (also known as 'partitioned CA' (Morita & Imai 1996a)) in which the bits that constitute one state are divided into some fields and treated separately. The Mode field stores the cell's current mode, the Link field does its direction, and the Gene field does the values of its internal registers, respectively.

State-transition rules

The state-transition rules used in this model are defined similarly to those of the original model by Morita and Imai, except for modifications made so as to keep the organisms working even in a situation of collision. These rules can be described in words as follows[3]:

- **For quiescent state:**
  - If stimulated by one of the adjacent active structure cells, it will turn into a blank structure cell rooted to the stimulating cell.
  - Otherwise it will remain quiescent.

- **For structure states:**
  - **Passive:**
    - If rooted to another structure cell, it will copy that cell's internal register values into its own, then become active if it received at least one gene and if there is no cell rooted to itself, i.e. it is the head of a pattern.
    - If not rooted to any structure cell, i.e. it is the tail of a pattern, it will encode which output interface is linked to other cells into its internal registers and become active.
    - In either case of the above two, if stimulated by one or more of the adjacent active structure cells, it will superimpose these cells' internal register values onto its own, using a logical 'OR' operation.
  - **Active:**
    - If rooted to another structure cell and there is no cell rooted to itself, i.e. it is the head of a pattern, it will become passive and copy the root cell's internal register values into its own. Then, if stimulated by one or more of the adjacent active structure cells, it will superimpose these cells' internal register values onto its own, using a logical 'OR' operation.
    - Otherwise it will become quiescent.

In the above description, 'rooted' means that the cell's input port is correctly linked to one of the output interfaces of another structure cell, and 'stimulated' means that there is an adjacent active structure cell that attempts to make a new structure cell onto that site according to direction by genes.

Note that there is no special rule provided here for cutting off the construction arm of self-replicating

[3]Contact me (sayama@necsi.org) for the complete rule set.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

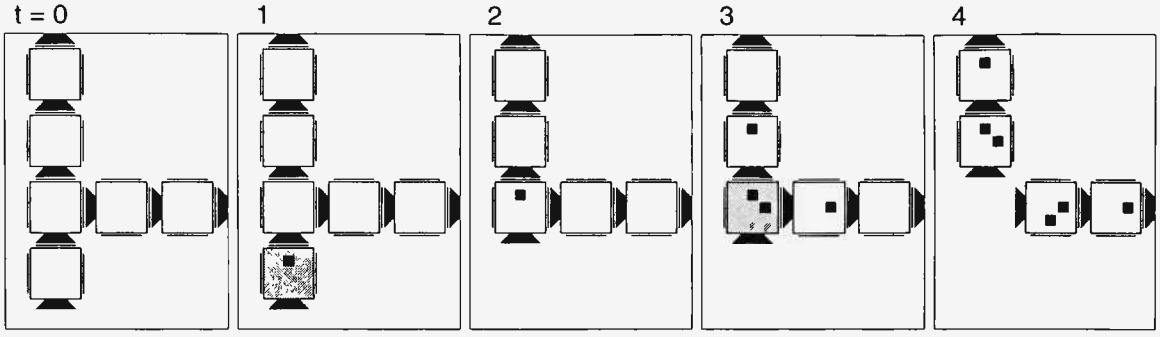

Figure 4: Shape encoding process at the tail of a pattern. The lowermost cell that has no root becomes active (indicated by gray in figures) and encodes which output interface is linked to other cells into its internal registers ($t = 1$). This active cell disappears while the encoded gene is conveyed upperward at the next time ($t = 2$). Such a process takes place repeatedly ($t = 3, 4$).

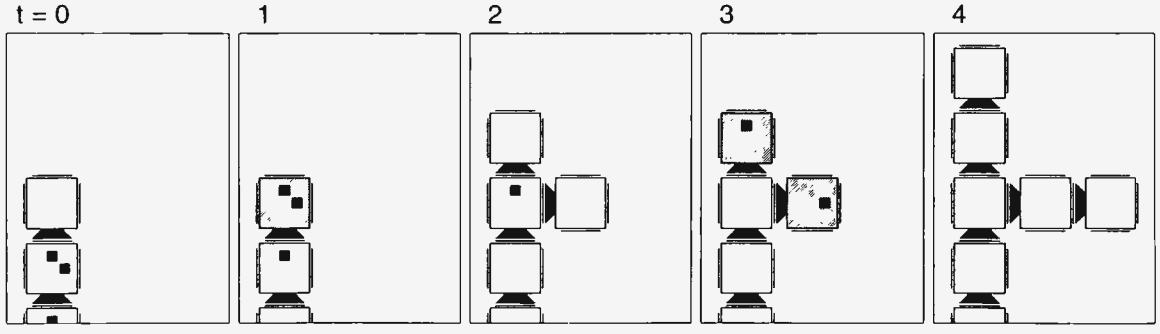

Figure 5: Shape decoding process at the head of a pattern. The uppermost cell to which no cell is rooted becomes active at the arrival of genes ($t = 1$). This active cell stimulates its adjacent quiescent sites to let them turn into blank structure cells according to direction by the genes ($t = 2$). Such a process takes place repeatedly ($t = 3, 4$).

loops that were often added to state-transition rules in earlier models. Since the primary motivation of this study is to extend the flexibility in shape of self-replicators, eliminating the constraint of loop structures is essential.

Behavior

Microlevel behavior

The shape-encoding/decoding behavior under the rules defined above are shown in Fig. 4 and 5. Shape encoding or decoding takes place in a cell which momentarily becomes active and determines the appearance of newly growing structure cells or its own disappearance.

The most interesting innovation introduced by this model, compared to earlier models, is that collisions of organisms lead to gene transmission beyond their boundaries, instead of irregular behavior or structural dissolution. This process is depicted in Fig. 6. Such interaction of phenotype, which directly affects the genotype, is not found for sophisticated life forms such as eukaryotic organisms including human beings. Bacteria Concerning the beginning of life, however, it may have been an important source of variation in driving evolution of primitive life forms born with so small complexity that there was no distinct line between genotype and phenotype.

Self-replication

A variety of patterns can replicate themselves in the proposed model due to the shape-encoding mechanism. One of the simplest self-replicating organisms in this model is shown in Fig. 7. It occupies only two sites in the space. It is remarkable that even such small organisms self-replicate through the interaction of genotype

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

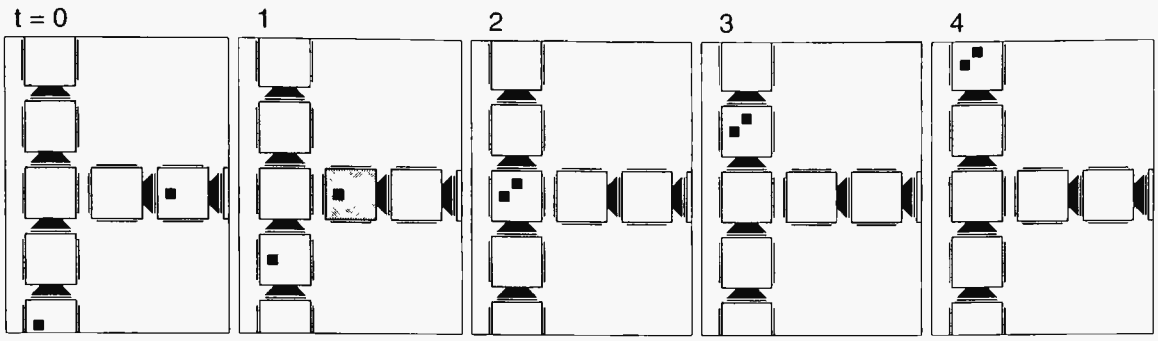

Figure 6: Gene transmission process occuring when a pattern collide into another. The central cell becomes active at the arrival of a gene ($t = 1$). However, the place onto which it wants to create a new structure cell is occupied by another existing pattern. Then the gene coming from the right pattern is transmitted beyond the boundary of two patterns and superimposed onto the left pattern's gene information using a logical 'OR' operation ($t = 2$).

and phenotype, involving transcription and translation of genes. Thus this process is non-trivial according to Langton's classification (Langton 1984). Other examples of small self-replicators are shown in Fig. 8.

Macrolevel behavior

This model displays interesting evolutionary behavior at the macro level. Simulations were conducted with initial configuration in which blank structure cells were randomly distributed at some specified density. All the results shown in this section are obtained using a square space of 200 × 200 sites with cutoff boundary conditions applied to its edges.

The typical outline of macrolevel behavior in this model is the following: (1) At first there is a short transient period when most of 'junk' patterns contained in the initial configuration are screened out of the space. (2) If some self-replicating organisms survive the initial transient period without being extinct, they begin to generate their respective colonies. (3) When the growing colonies crash against each other (or against sterile patterns remaining), in some cases one of the competing clusters absorbs the other; in other cases the crash happens to generate a new kind of cluster and it overcomes its 'parent' clusters. (4) Eventually, the space falls into one of the following types of final states:

Type I Static or periodic state with no self-replicator.

Type II Dynamic state in which the space is filled like a mosaic with clusters each of which is made of self-replicators of the same kind.

Type III Dynamic state in which the space is filled with a dense cloud made of a jumble of complicated self-replicators with low self-replication fidelity.

Type IV Static (or almost static) state in which the space is filled with infinitely growing static webs that originated in looped patterns included in the initial configuration.

Examples of these four final states are shown in Fig. 9. Which type the system finally falls into is dependent on the number and the kind of self-replicating patterns that survive the initial transient period. In general, larger space and higher initial density of structure cells make the initial population more diverse, which leads to the greater probability of the appearance of type III or IV final state. How the final states depend on initial density of structure cells is roughly shown in Fig. 10.

The most interesting behavior is the process of evolution toward the type III final state, which is principal behavior of this model for a regime with the initial density between 0.15 and 0.3. Fig. 11 shows a typical example. In this case, at first two colonies of simple self-replicators are formed after initial screening period (t = 0–108). The collisions among them and other sterile patterns give rise to appearance of some other self-replicators (t = 330). Once a dense cloud of more complicated self-replicators is formed (t = 500), it gradually proliferates (t = 760–1000) and finally fills up all the space (t = 1500–2000). This process looks like an evolutionary process in real biology performed by variation and natural selection. However, we should note that, since the organisms in this model change their genotype or phenotype through direct interaction very frequently, it is no longer possible to trace their lineages and consider their reproduction and selection dynamics to be the same as that in real biology.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

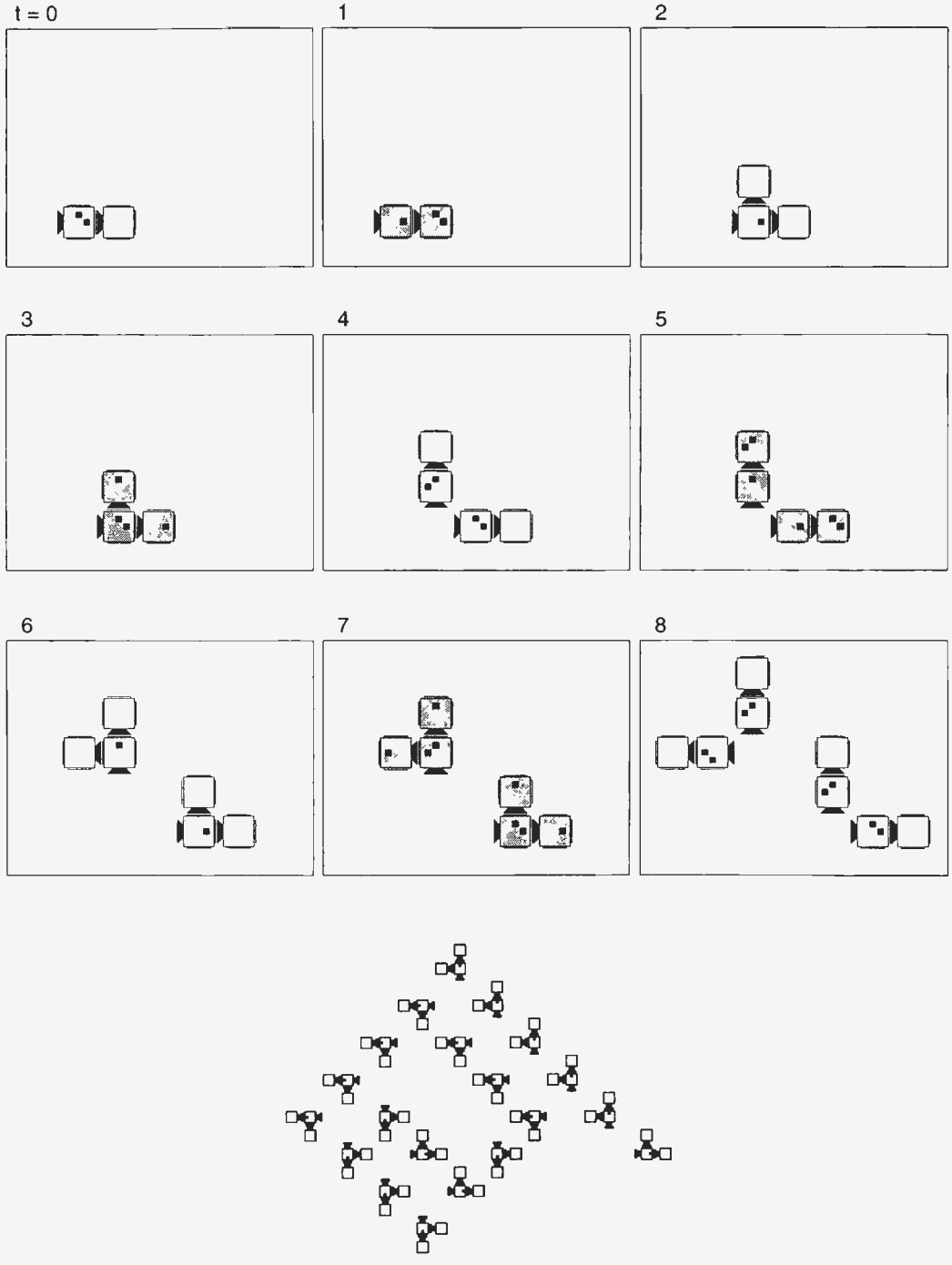

Figure 7: One of the simplest self-replicating organisms. This worm travels straight and emits its offspring to the left side repeatedly. The lower figure is a growing colony of them.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

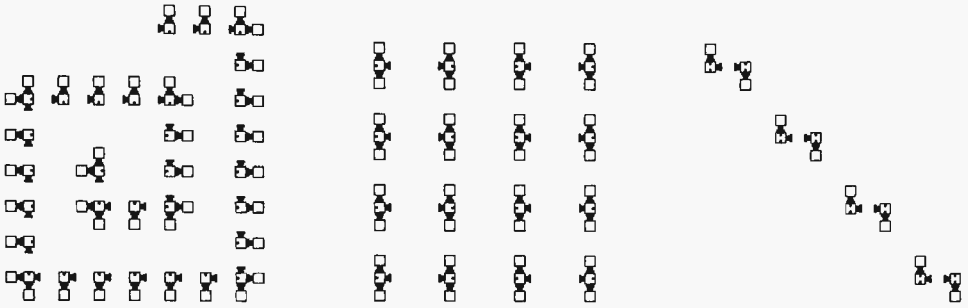

Figure 8: Examples of other kinds of simple self-replicators.

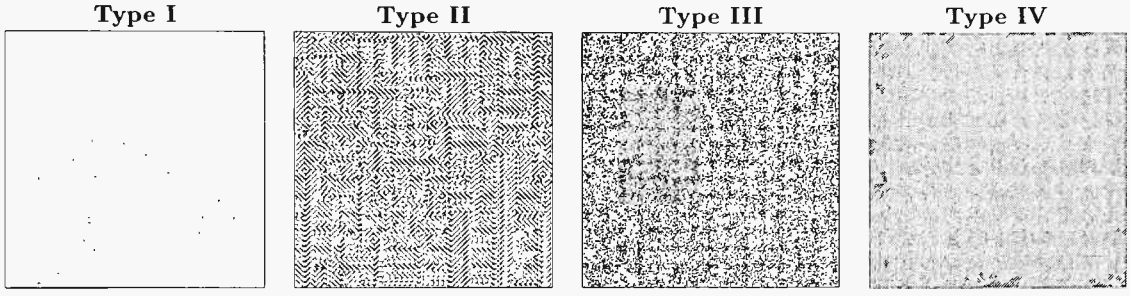

Figure 9: Four types of the final states of evolution. For simplicity, only the direction of structure cells is plotted. In the initial configurations are randomly directed structure cells set onto 10% of all the sites for type I, II, III, and 30% for type IV. Type I: At $t = 200$ with random number seed 1234. Type II: At $t = 1500$ with random number seed 4321. Type III: At $t = 2000$ with random number seed 12345. Type IV: At $t = 1000$ with random number seed 1234. (With these seed numbers you can simulate them again using a Java applet intoduced at the end of this article.)

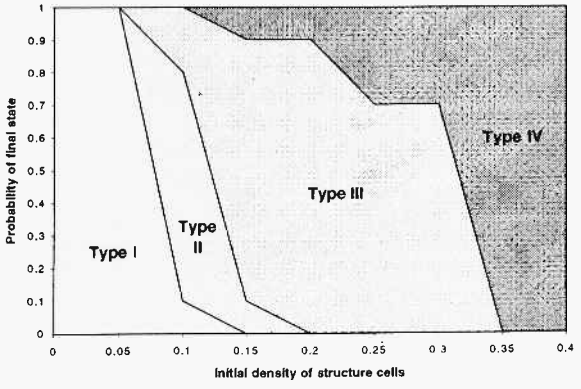

Figure 10: Dependence of final states upon initial density of structure cells. Ten simulation results are shown for each initial density. The space is of 200×200 sites.

Increase of structural complexity

In the evolution toward the type III final state, a kind of increase of structural complexity is observed. Temporal development of structural complexity of the organisms in Fig. 11 is characterized in Fig. 12 using the average of length of organisms, the average frequency of branching per organism, and the average number of genes per organism. These graphs show that all the measured quantities are increasing after the appearance of the dense cloud of complicated worms around $t = 500$. The emergence and fixation of more complicated worms implies that this model has a nature to favor those with such complicated shape to some extent.

However, such increase of complexity always saturates at some constant level, i.e. the evolution in this model is definitely restricted. Fig. 13 shows an example of such a saturated situation taken from the final state in Fig. 11. It is observed in this figure that each

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

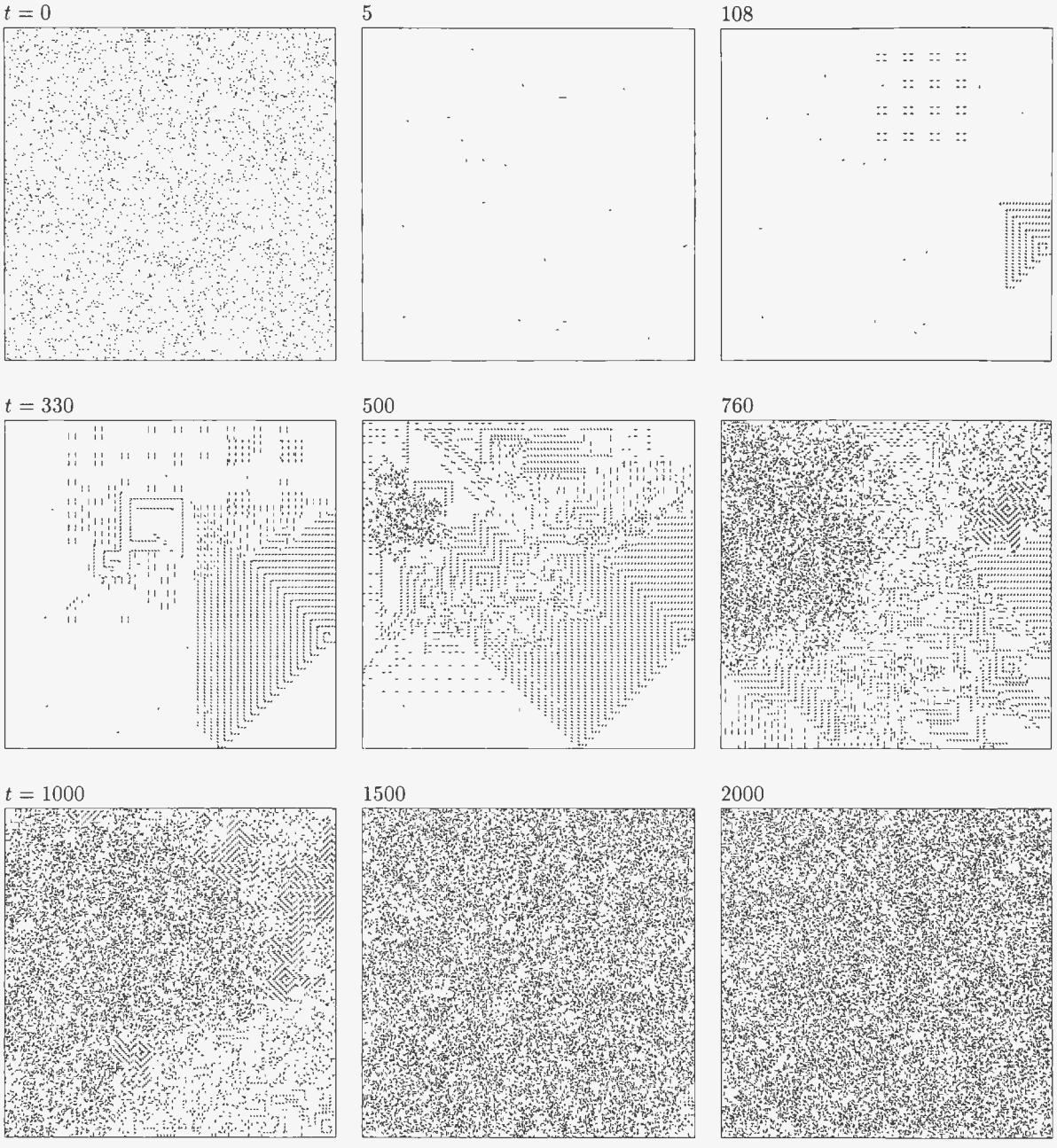

Figure 11: Evolution of worms from random initial configuration with 10% sites of structure cells. For simplicity, only the direction of structure cells is plotted. The random number seed used is 12345.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

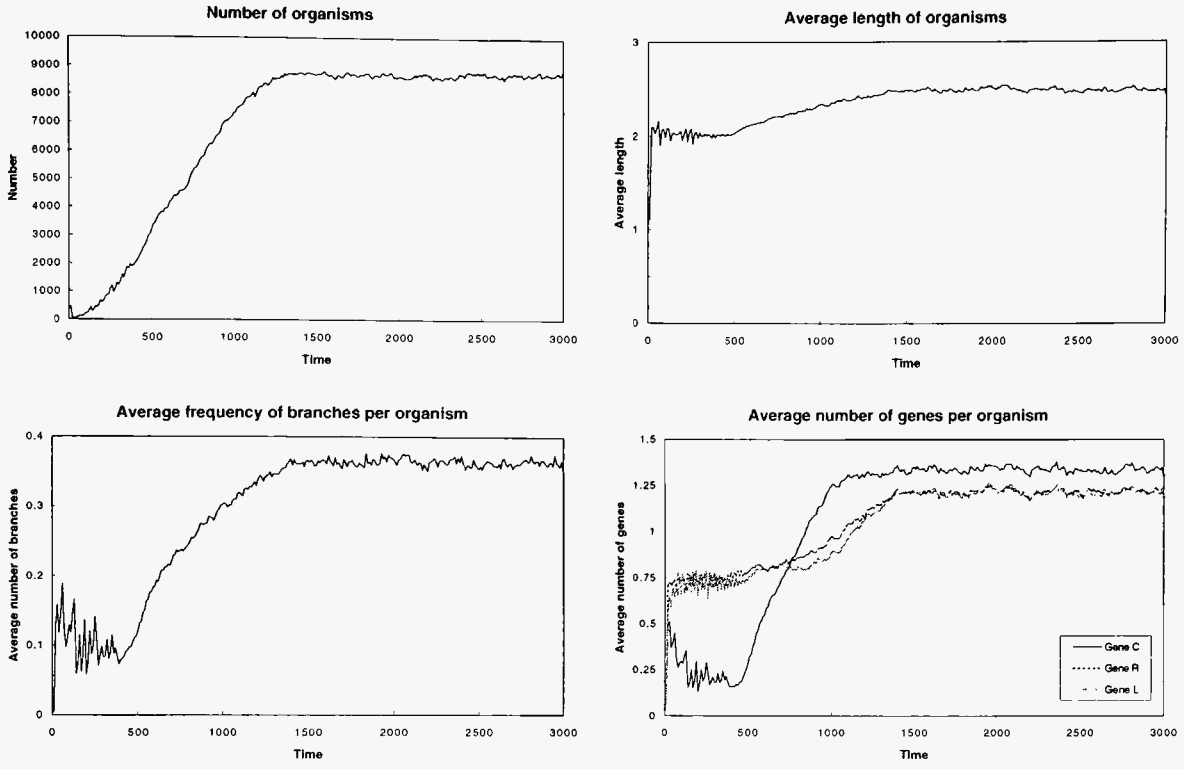

Figure 12: Evolution of structural complexity characterized by the average length of organisms (upper right), the average frequency of branching per organism (lower left), and the average number of genes per organism (lower right) in the case shown in Fig. 11.

organism has many gene information accumulated but most of them are lost without being translated to phenotype. The self-replication fidelity is thus diminishing there due to overcrowding, where the shape of offspring is determined more significantly by the environmental constraints, i.e. the availability of room around the organism for its growth.

One of the possible reasons for this saturation is the shortage of local room necessary for full translation from genotype to phenotype due to the fixation of place and shape of organisms onto particular sites. We find in Fig. 13 that there are still some areas of unused sites left near the crowded organisms. They cannot make use of such empty areas in their vicinity because their location and shape are strictly fixed to particular sites, which is one of the intrinsic features of CA. In contrast, biochemical polymers in real cells such as DNA can move and change shape adaptively reacting upon external forces, which enables many ribosomes simultaneously translate genetic information into proteins in a very compact local area. To realize open-ended evolution on CA-like artificial media, it would be necessary to develop and use a newer model of space that has some fluidity in terms of place and shape of virtual organisms.

Conclusion

In this article, I introduced a new self-replication model with shape-encoding mechanism on a sixty-five-state CA space with von Neumann neighborhood, where virtual organisms transmit genetic information over their boundaries through collisions of phenotypes. An interesting result was obtained from simulations that there were some cases where the structural complexity of organisms characterized by the average length of organisms and the average frequency of branching per organism increased as the population evolved. However, an unlimited increase of structural complexity could not occur in this model. This limitation should be caused in part by an intrinsic problem of CA that the location and the shape of virtual organisms are strictly stuck to particular sites so that there is no local room

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022

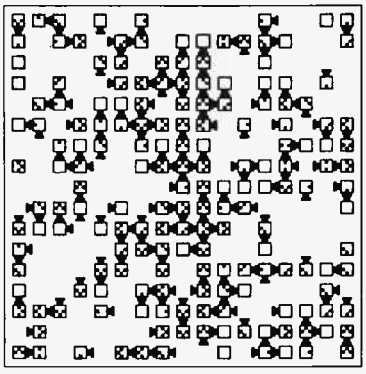

Figure 13: Enlargement of a central part of the saturated population in Fig. 11 at $t = 2000$.

to allow further complicated structures to evolve. It would be necessary to introduce some fluidity of site arrangements into the model for further development of artificial evolutionary models using CA-like space.

For the readers who may want to see the dynamic behavior of the worms introduced, a Java applet for simulating them was developed and is available at `http://necsi.org/postdocs/sayama/worms/`.

Acknowledgments

I would like to thank Kenichi Morita and Katsunobu Imai for permission to use their figure in this article. Thanks are also due to Yaneer Bar-Yam, Barry McMullin, and two anonymous reviewers for valuable comments. This research was supported in part by a grant from the Murata Overseas Scholarship Foundation.

References

Byl, J. 1989. Self-reproduction in small cellular automata. *Physica D* 34:295–299.

Chou, H. H., and Reggia, J. A. 1997. Emergence of self-replicating structures in a cellular automata space. *Physica D* 110:252–276.

Chou, H. H., and Reggia, J. A. 1998. Problem solving during artificial selection of self-replicating loops. *Physica D* 115:293–312.

Codd, E. F. 1968. *Cellular Automata.* ACM Monograph Series. New York: Academic Press.

Laing, R. 1977. Automaton models of reproduction by self-inspection. *Journal of Theoretical Biology* 66:437–456.

Langton, C. G. 1984. Self-reproduction in cellular automata. *Physica D* 10:135–144.

Lohn, J. D., and Reggia, J. A. 1995. Discovery of self-replicating structures using a genetic algorithm. In *Proceedings of the 1995 IEEE International Conference on Evolutionary Computation (ICEC'95)*, 678–683. Piscataway, New Jersey: IEEE.

Lohn, J. D., and Reggia, J. A. 1997. Automatic discovery of self-replicating structures in cellular automata. *IEEE Transactions on Evolutionary Computation* 1(3):165–178.

Marchal, P. 1998. John von Neumann: The founding father of artificial life. *Artificial Life* 4(3):229–235.

Morita, K., and Imai, K. 1996a. Self-reproduction in a reversible cellular space. *Theoretical Computer Science* 168:337–366.

Morita, K., and Imai, K. 1996b. A simple self-reproducing cellular automaton with shape-encoding mechanism. In Langton, C. G., and Shimohara, K., eds., *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, 489–496. Nara, Japan: MIT Press.

Perrier, J. Y.; Sipper, M.; and Zahnd, J. 1996. Toward a viable, self-reproducing universal computer. *Physica D* 97:335–352.

Pesavento, U. 1995. An implementation of von Neumann's self-reproducing machine. *Artificial Life* 2(4):337–354.

Reggia, J. A.; Armentrout, S. L.; Chou, H. H.; and Peng, Y. 1993. Simple systems that exhibit self-directed replication. *Science* 259:1282–1287.

Sayama, H. 1998a. *Constructing evolutionary systems on a simple deterministic cellular automata space.* Ph.D. dissertation, Department of Information Science, Graduate School of Science, University of Tokyo.

Sayama, H. 1998b. Introduction of structural dissolution into Langton's self-reproducing loop. In Adami, C.; Belew, R. K.; Kitano, H.; and Taylor, C. E., eds., *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*, 114–122. Los Angeles, California: MIT Press.

Sayama, H. 2000. A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life* 5(4):343–365.

Sipper, M. 1994. Non-uniform cellular automata: Evolution in rule space and formation of complex structures. In Brooks, R., and Maes, P., eds., *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, 394–399. Cambridge, Massachusetts: MIT Press.

Sipper, M. 1998. Fifty years of research on self-replication: An overview. *Artificial Life* 4(3):237–257.

Tempesti, G. 1995. A new self-reproducing cellular automaton capable of construction and computation. In Morán, F.; Moreno, A.; Merelo, J. J.; and Chacón, P., eds., *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life (ECAL'95)*, volume 929, Lecture Notes in Artificial Intelligence of *Lecture Notes in Computer Science*, 555–563. Granada, Spain: Springer-Verlag.

Vitányi, P. M. B. 1973. Sexually reproducing cellular automata. *Mathematical Biosciences* 18:23–54.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata.* Urbana, Illinois: University of Illinois Press. edited and completed by A. W. Burks.

Copyrighted Material

Downloaded from http://direct.mit.edu/books/book/chapter-pdf/268063/9780262291071_cac.pdf by guest on 11 August 2022