# Self-Evolution in a Constructive Binary String System

Peter Dittrich[1] and Wolfgang Banzhaf[2]

Dept. of Computer Science, University of Dortmund [3]
D-44221 Dortmund (Germany)

- DRAFT January 12, 1998-

**Abstract**

We examine the qualitative dynamics of a catalytic self-organizing reaction system of binary strings that is inspired by the chemical information processing metaphor. A string is interpreted in two different ways: either (i) as raw data or (ii) as a machine which is able to process another string as data in order to produce a third one. This paper focuses on the phenomena of evolution whose appearance is notable because no explicit mutation, recombination or artificial selection operators are introduced. We call the system self-evolving because every variation is performed by the objects themselves in their machine form.

Keywords: self-organisation, artificial life, prebiotic evolution, molecular computer, self-programming

[1] e-mail: dittrich@@LS11.informatik.uni-dortmund.de
[2] e-mail: banzhaf@@LS11.informatik.uni-dortmund.de
[3] URL: http://LS11-www.informatik.uni-dortmund.de

# 1 Introduction

In recent years computer simulations have been used to study the phenomena of life, not by simulating life as-it-is (weak AL) but by instantiating life as-it-could be (strong AL) [?]. AL systems have been used, e.g. for studying questions concerning the emergence and quality of organizations [?, ?, ?], the process of diversification [?], the origin of replicators [?], or morphological evolution [?].

An important property of most strong AL systems is that they contain the ability for self-reference. For instance Ray's Tierra organisms are able to read, copy and modify their own code. In Fontana's Algorithmic Chemistry every object is a character string able to process other objects by using the lambda-calculus which maps the character string into an (active) function. The dualism inherent in those systems can be traced back to Goedel [?] who defined a mapping of mathematical statements into natural numbers that allowed self-reference and to von Neumann's stored program computer [?].

The system discussed here is inspired by a chemical reaction dynamics, where molecules collide and interact to create new molecules forming metabolic networks. This contribution tries to give insight into the origin and evolution of these networks [?]. Related work [?, ?, ?, ?, ?] focuses on the static structure of the emerged networks. Here we will concentrate on dynamic phenomena, especially on the emergence of prebiotic evolution [?]. Evolutionary phenomena can be observed, although no explicit fitness, mutation, recombination or selection operators are used. Variations are only performed by the objects (molecules, binary strings) themselves when they act in machine form.

## 1.1 Self-organization and Evolution

It seems worthwile to start our considerations with a short note on self-organization and what we mean by "evolution" in the context of this paper. **Self-organization** is a process where a system is organized while the components directing this process are part of the system. This kind of view of "self-organization" implies: 1.) Every organization phenomenon becomes a result of self-organization provided we enlarge the system boundaries. 2.) Self-organization need not to be directed.

When transforming this linguistic abstract description of self-organization into the language of computer science it seems to be straight forward to represent an organizer as a program and the data a program is working on as the target for organization. In order to close the loop inherent in self-organization we need a mapping from data to machines or vice versa, yielding the dualism of the system components. The information residing in the system should either act as *active* machines or should be processed as *passive* data.

The organization process can also be expressed in the terms of mathematics as the application of (active) operators to (passive) data. An example for this approach is [?] where the operands are binary strings. Operators are formed out of binary strings by "folding" them into a matrix.

The outline of this paper is as follows: Section 2 gives a general introduction to algorithmic reaction systems. In Section 3 we briefly describe the special reaction mechanisms used for this contribution. Section 4 summarizes methods for investigation and visualisation. In Section 5 we discuss four qualitative types of dynamic self-oranizattion phenomena by showing

typical simulation results, namely (1) extinction, (2) emergence of an oranization structure, (3) exploration/innovation, and (4) evolution.

## 2 Algorithmic Reaction Systems

The system used for the experiments in this paper is called an **algorithmic reaction system**, because the interaction between the participants is specified by an algorithm. It consists of the following three components:

- *A soup (population) of objects.*
  These objects may be abstract symbols [?], character sequences [?], lambda-expressions [?], binary strings [?, ?], numbers [?] or proofs [?]. Here, we use binary strings with a constant length of 32 bit. In a basic setting, the soup has no topological structure so that its state can be noted as a concentration vector.

- *A collision or reaction rule.*
  The collision rule defines the interaction among two objects $s_1$ and $s_2$ which may lead to the generation of a new object $s_3$. This is denoted as:

  $$s_1 + s_2 \Longrightarrow s_3$$

  Note that the operator "+" is *not* commutative.

- *An algorithm to run the system.*
  In this contribution we use an algorithm that can also be found with minor modifications in [?, ?, ?, ?].

**Reactor algorithm**

1. Select two objects $s_1, s_2$ from the soup randomly, without removing them.
2. If there exists a reaction $s_1 + s_2 \Longrightarrow s_3$ and the filter condition $f(s_1, s_2, s_3)$ holds, replace a randomly selected object of the soup by $s_3$.

The replaced objects form the **dilution flux** of the system. The **filter** $f$ can be used to block lethal objects and to introduce elastic collisions easily. The term **collision** refers to one execution of the two steps of the reactor algorithm. A collision is called **elastic** if no product $s_3$ is inserted into the soup. So, a collision of two objects $s_1, s_2$ is elastic, if no product if defined by the reaction rule or if the filter condition prohibits the insertion of the product. An object is said to be **lethal** if it is able to replicate in an unproportionally large number in almost any ensemble configuration.

The interaction scheme of the algorithm can be written as a chemical reaction equation:

$$s_1 + s_2 + X \longrightarrow s_1 + s_2 + \begin{cases} \frac{1}{2}s_3 & \text{with } s_1 + s_2 \Longrightarrow s_3 \\ \frac{1}{2}s_3' & \text{with } s_2 + s_1 \Longrightarrow s_3' \end{cases}$$

In other words, that $s_1$ and $s_2$ are not consumed but act as catalysts of the reaction. The **raw material** $X$ is used to balance the equation. It does not appear explicitly in the system and

could be interpreted as computational resources like processing time or memory [?]. Reaction systems exhibiting this kind of reaction scheme are able to form **hypercyclic** oranizations [?].

In order to measure the running time of the algorithm we call $M$ iterations (collisions) a **generation**, where $M$ is the soup or reactor size. Using "generations" rather than "number of iterations" allows an easier comparison of runs with different soup sizes.

### 2.1 A model for the algorithmic reactor

In a setting corresponding to a well-stirred tank reactor the state of the system can be described by a concentration vector $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_1 + \cdots + x_n = 1$ and $x_i > 0$. $x_i$ is the concentration of string type $s_i$. For a large and constant population size the behavior of the reactor algorithm is modeled by the **catalytic network equation** [?]:

$$\boxed{\frac{dx_k}{dt} = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij}^k x_i x_j - x_k \sum_{i,j,k=1}^{n} \alpha_{ij}^k x_i x_j \qquad k = 1, \ldots, n} \tag{1}$$

with second order rate constant $\alpha_{ij}^k$ for the reaction $i + j \overset{\alpha_{ij}^k}{\Longrightarrow} k$ .

Here the rate constants become

$$\alpha_{ij}^k = \begin{cases} 1 & \text{if } s_1 + s_2 \Longrightarrow s_3 \text{ and the filter condition } f(s_1, s_2, s_3) \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

Equation 1 becomes the famous **replicator equation** if for all $i, j \in \{1, \ldots, n\}$

$$\alpha_{ij}^k = 1 \Longrightarrow i = k \vee j = k$$

holds [?, ?].

Simulating systems with a large number $n$ of different objects becomes difficult. For the special case if only a small fraction of all possible objects being present in the reactor, a technique called **meta dynamics** is used elsewhere, where the ODE Sytem (equation 1) is treated dynamically and updated under certain conditions. For example equations are removed when the corresponding concentration falls below a given theshold [?, ?] . The advantage of this method is, that a potentially infinite number of objects can be handled by still using the ODE framework. Problems arise, however, when the diversity (e.g. the number of different objects in the reactor) becomes high, as will be the case here.

## 3 Special 32-bit Reaction Mechanisms

We will now briefly describe the special reaction mechanisms used for this contribution, namely the AND reaction and the automata reaction.

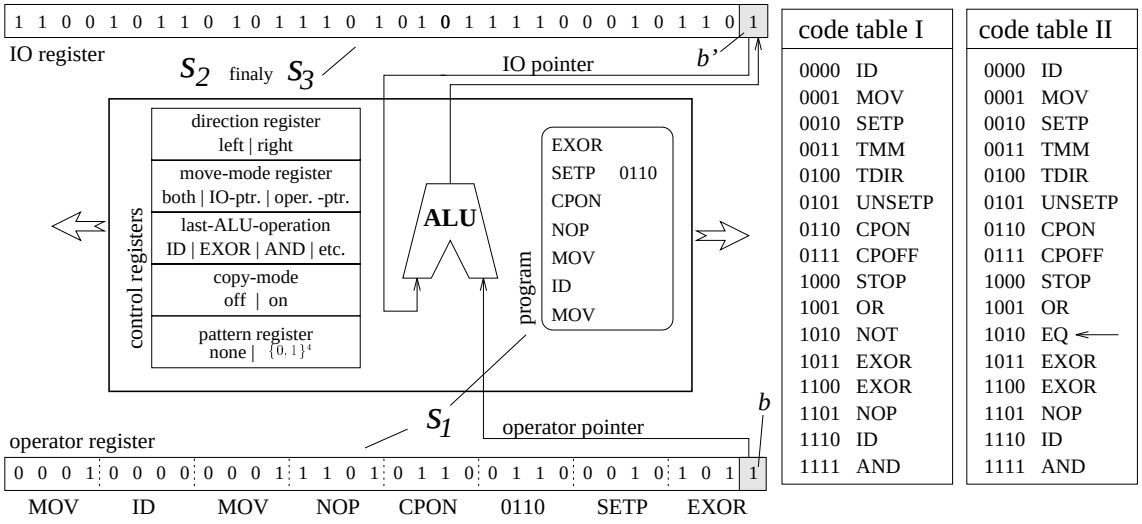



Figure 1: *Automaton, resulting by folding $s_1$. It carries out the reaction $s_1 + s_2 \Longrightarrow s_3$. $s_1$ is written into the operator register and specifies the program. The IO register is initialized with $s_2$ and contains the result $s_3$ after running the program.*

## 3.1 AND reaction

The AND-Reaction is simply the bitwise logic AND of strings $s_1$ and $s_2$.

$$\forall i \in \{0, 31\} : s_3^{(i)} = s_1^{(i)} \wedge s_2^{(i)}$$

where $s^{(i)} \in \{0, 1\}$ denotes the $i$-th bit of string $s$. This reaction is used for test and reference, because its behavior can be easily described and understood.

## 3.2 Automata reaction

The **automata reaction** is based on a finite state automaton which is a mixture of a Turing-machine and a register machine. It has also been inspired by the Typogenetics of Hofstadter [**?**].

The **automata reaction** instantiates a deterministic reaction $s_1 + s_2 \Longrightarrow s_3$, where $s_1, s_2, s_3 \in \{0, 1\}^{32}$. In order to calculate the product $s_3$, string $s_1$ is folded into an automaton $A_{s_1}$, which gets $s_2$ as an input. The construction of $A_{s_1}$ ensures that the automaton will halt after a bounded finite number of steps. Because $A_{s_1}$ is a deterministic finite automaton, the automata reaction defines a functions $\{0, 1\}^{32} \times \{0, 1\}^{32} \longrightarrow \{0, 1\}^{32}$.

Figure 2 shows the structure of the automaton. It contains two 32-bit registers, the **IO register** and the **operator register**. At the beginning operator string $s_1$ is written into the operator register and operand $s_2$ into the IO register. The program is generated from $s_1$ by simply mapping successive 4-bit segments into instructions. Note that two slightly different mappings are used (Figure 2 , right). The resulting program is executed sequentially, starting with the first instruction. There are no control statements for loops or jumps in the instruction set.

Each 32-bit register has a pointer, refering to a bit location. The **IO pointer**, refering to a bit $b'$ in the IO register and the **operator pointer**, refering to a bit $b$ in the operator register. Bit

| $s_1$ | command | comment |
|---|---|---|
| 1011 | EXOR | The exclusive-or product of $b'$ (refered to by the IO pointer) and $b$ (refered to by the operator pointer) is written at the position of $b'$. Therefore, the 0. bit of the IO register becomes 0. Then both pointers are moved one bit position to the left. |
| 0010 0110 | SETP 0110 | The pattern register is set to 0110. If the pattern is set, a subsequent MOV operation will move the pointers until the pattern is found in the operator register (max. 32 steps). |
| 0110 | CPON | The copy-mode is switched on. Every time the pointers are moved the ALU operation is performed (here, EXOR). |
| 1101 | NOP | No operation. The registers are not modified. |
| 0001 | MOV | The pointers are moved to the left until the pattern 0110 is found in the operator register, which happens at the 8. bit position. Because the copy-mode is activated (by CPON), the last ALU operation (here: EXOR) is executed for every bit passed. |
| 0000 | ID | The 8. bit of the operator register is copied into the IO register, and both pointers are moved one step to the left. |
| 0001 | MOV | Both pointer are moved to the next apearance of 0110, namely to the 12. bit position. This time the ID operation is executed at each step. |

Table 1: Example for the automata reaction.

$b$ and $b'$ are inputs to the **ALU**. The ALU result is stored at the IO pointers location, therefore replacing $b'$.

Instead of going into more details now, an example will be presented. The Appendix contains descriptions of the register configuration and the instructions.

### 3.2.1 Example for the automata reaction

Table 1 shows the function of the automaton for the following example:

$$s_1 = 0001\,0000\,0001\,1101\,0110\,0110\,0010\,1011$$
$$s_2 = 1100\,1011\,0101\,1101\,0101\,1110\,0010\,1101$$
$$s_3 = 1100\,1011\,0101\,1101\,0101\,0110\,0000\,0110$$

In hex notation, this reads:

$$101d662b + cb5d5c2d \Longrightarrow cb5d5606$$

## 4 Investigation and Visualisation of System Behavior

Designing a system that shows behavior of enormous complexity is surprisingly easy. But to visualize and to explain in detail the behavior is much harder. The methods for investigation can be divided into macroscopic, mesoscopic and microscopic ones.

A **macroscopic method** observes the whole system, by mapping its high-dimensional state into a low-dimensional space, where every component of the system should be considered. An

example from thermodynamics is the temperature. A **mesoscopic method** considers only a small subset of the systems components. It still accumulates many microscopic events into low dimensional values (e.g. the concentration of specific substance in a region). **Microscopic methods** are visualizing elementary details of the system for instance the history of a specific object or memory cell.

Here, we shall apply only macroscopic methods, which are explained in the following:

**Diversity**: Various measure for the diversity of a population exist. Most methods require a grouping $G$ of the individuals.

$$G : P \to \mathbb{N}, \quad \mathrm{G(s_i) = g_i}$$

where $P = \{s_1, \ldots, s_M\}$ is a population of genomes and $g_i$ is the number of the group $s_i$ belongs to. In ecology, diversity is often defined as the number of different groups or taxa:

$$Div_1(P) = |\{g | \exists s \in P : G(s) = g\}| \tag{2}$$

The dynamic behavior of $Div_1$ depends on the definition of the grouping $G$. A fine grained grouping results in a diversity fluctuating over time. This is due to the fact that the number of individuals belonging to one group is not considered in equation 2.

Here, we use a simple grouping, where every possible genotype forms a group. The resulting (absolute) diversity $Div_{abs}$ then becomes the number of different strings in $P$.

$$Div_{abs}(P) = |\{s | s \in P\}|$$

To get a relative value $Div(P)$ in the range $0 - 1$ the absolute diversity is divided by the soup size $M$. In the future, the term "diversity" refers to the **relative diversity**

$$\boxed{Div(P) = \frac{Div_{abs}(P)}{M}.}$$

**Distance distribution complexity** (**DCC**): One problem of the above diversity measure is that the distance between different groups is not taken into account. For example an ecology with 10 horses and 10 zebras would have the same diversity as an ecology with 10 horses and 10 dolphins. The DDC, recently introduced by Kim [**?**], is able to distinguish these cases. Given a discrete distance measure $D : P \times P \to \mathbb{N}$ the **distance distribution** is defined here as the relative frequency of the distance value $d \in \mathbb{N}$:

$$f(d) = \frac{2}{M(M-1)} |\{(s_i, s_j) | s_i, s_j \in P, i < j, D(s_i, s_j) = d\}|$$

[**?**]. The **distance distribution complexity** (DDC) is then defined as the Shannon entropy of the distribution of distance values:

$$\boxed{DDC(P) := -\sum_d f(d) \log(f(d))}$$

For the following analysis $D$ is defined as the Hamming-distance. This simple and fast measure is sufficient, because no shifting and no insertion or deletion of bits is taking place. Note, a high

DDC indicates a high diversity and vice versa, but the DDC can increase even if the diversity decreases (e.g. Figure 3 ).

**Productivity**: The **productivity** is the probability, that a collision of two strings is reactive, thus

$$\boxed{Prod(P) = p(\text{a string is inserted into the soup in the next iteration of the reactor algorithm}).}$$

If the productivity is zero no new string is produced. To measure the productivity during a simulation, the number of iterations in which a string $s_3$ is inserted into the soup is counted for $M$ steps. The result is divided by $M$. If the soup size $M$ is constant the productivity is equal to the dilution flux.

**Innovativity**: The innovativity is the probability that a collision produces an object that is new, with respect to a given time window $[t_1, t_2]$. Let $P_t$ be the soup after the $t$-th iteration of the reactor algorithm. Then the **(total) innovativity** is defined for the time window $[0, t]$ as

$$\boxed{Inn(P_t) = p(s_3 \text{ is inserted into the soup} \quad \text{and} \quad s_3 \notin \bigcup_{t'=0}^{t} P_{t'})}$$

The total innovativity has been measured by counting the number of completly new strings that are inserted into the soup during $M$ iterations of the reactor algorithm. The result is divided by $M$. There are more local measures of innovativity which consider a smaller time window of constant size.

**Reaction table**: Given a set of objects $A = \{s_1, \ldots, n_a\}$ then the corresponding reaction table $RT : \{1, \ldots, n_a\}^2 \to \{1, \ldots, n_a, -, *\}$ is defined as:

$$RT(i,j) = \begin{cases} - & \text{if the collision of } s_i, s_j \text{ is elastic,} \\ * & \text{if } s_i + s_j => s_k \text{ and } s_k \notin A, \\ k & \text{if } s_i + s_j => s_k \text{ and } s_k \in A. \end{cases}$$

The term "the collision of $s_i, s_j$ is elastic" refers to step 2 of the reactor algorithm, where $s_i, s_j$ are selected but no product is inserted into the soup.

For the reaction tables shown in this paper, the set of objects is sorted according to their concentration:

$$\forall s_i, s_j \in A : [s_i] \geq [s_j] \Longleftrightarrow i \geq j$$

# 5 Qualitative Types of Dynamic Self-Organization Phenomena

In this section we show some experimental results to discuss certain qualitative types of self-oranization processes. In all experiments the objects are binary strings of fixed size 32 bit. The soup is always initialized with random strings, resulting in a initial diversity of approximately 1.

The types will be illustrated by visualizing "typical" runs. If the behavior of a system is simple, a series of runs with the same parameter but different random numbers can be summarized in

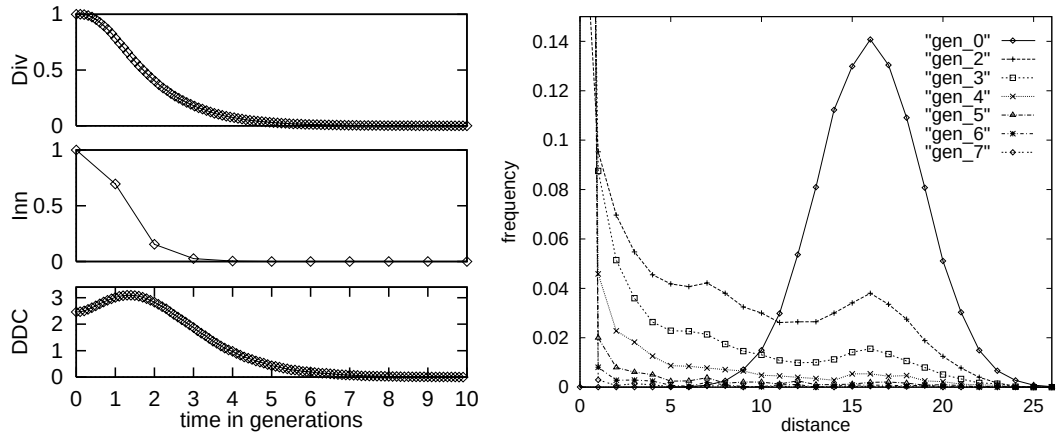



Figure 2: *Example for extinction. Paramters: AND reaction, no filter condition,* $M = 10^4$. *The diversity and the DDC drop to zero because the system is exploited by the destructor* 00000000 *which remains as the only surviving string. Right: Distance distribution for the generations 0-7.*

one diagramm, by simply forming the avarage of the displayed values. But this makes very little sense if the behavior of the system is complex. It is problematic, because experimentators subjective ability of image processing is used to compare the runs.

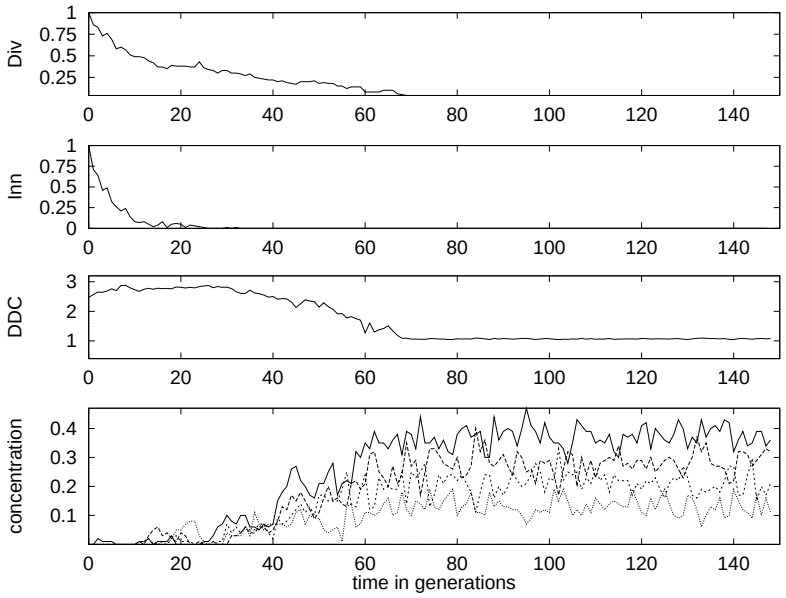
## 5.1 Extinction

Figure 3 shows the behavior of the AND reaction. The diversity drops quickly to $1/M$ (only one string type left) and the DDC to zero, indicating that the soup is exploited by a single type, the lethal string 00000000. It is able to replicate with every other string and is furthermore produced through many reaction pathways (e.g. $f120000 + 0004711a \Longrightarrow 00000000$).

At the outset the DDC increases, because intermediate products appear which are characterized by a small number of ones (e.g. 00000200), resulting in a higher complexity of the distribution of distances during generation 0-2 (Figure 3 , right).

## 5.2 Emergence of an organization

Figure 4 shows the emergence of a simple organization comprising four different string types. The diversity and innovativity drop continuously until a closed self-maintaining [4] organization dominates the soup with only a few distinct objects. Through genetic drift [?] the diversity of this organization is slowly reduced until an organization remains, which does not contain a closed, self-maintaining subset.

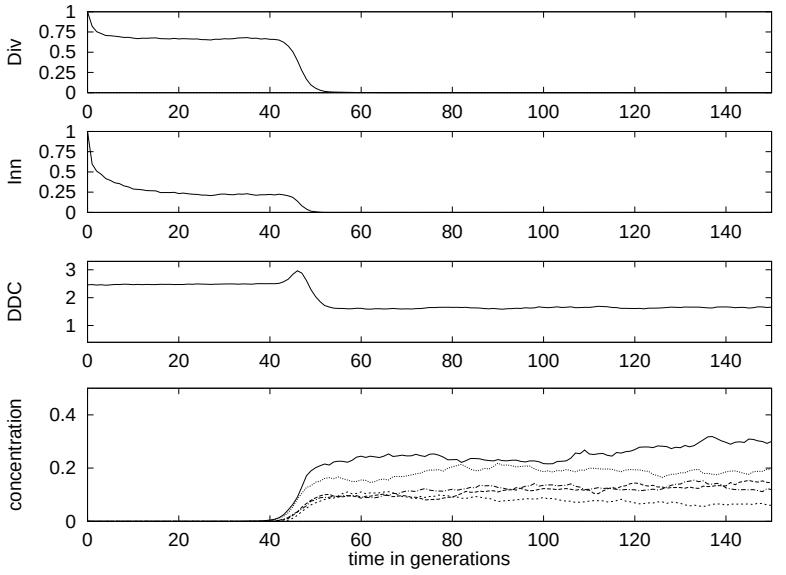



Reaction table at generation 140

| No | String | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| 0 | 7240a7ef | 2 | 1 | 2 | 1 |
| 1 | 7240a7ea | 0 | 0 | 0 | 0 |
| 2 | 7240a7eb | 3 | 0 | 3 | 0 |
| 3 | 7240a7ee | 1 | 1 | 1 | 1 |

| String | Program |
|---|---|
| 1111 | AND |
| 1110 | ID |
| 0111 | CPOFF |
| 1010 | NOT |
| 0000 | ID |
| 0100 | TDIR |
| 0010 | SETP 0111 |
| 0111 | |

Figure 3: *Emergence of an organization. Parameters: Automata reaction with code table 1, $M = 100$. no filter condition, soup seeded with $M$ random strings. For the four surviving strings the reaction table and their concentration over time is given.*




Reaction table at generation 280

| No | String | 0 | 1 | 2 | | 7 |
|---|---|---|---|---|---|---|
| 0 | 1e1ca260 | 0 | 0 | 0 | | 0 |
| 1 | 1e1ca261 | 1 | 1 | 1 | | 1 |
| 2 | 1e1ca264 | 2 | 2 | 2 | | 2 |
| | | | | | | |
| 7 | 1011a261 | 7 | 7 | 7 | | 7 |

| String | Program |
|---|---|
| 0000 | ID |
| 0100 | CPON |
| 0010 | SETP 1010 |
| 1010 | |
| 1100 | EXOR |
| 0001 | MOV |
| 1110 | ID |
| 0001 | MOV |

Figure 4: *Exploration and innovation. During the explorative phase (generation 2 - 40) the diversity is constantly high and a lot of totaly new strings are produced. Parameters: soup size $M = 10^4$, automata reaction with table 1, no filter condition $f$, soup seeded with $M$ random strings.*

## 5.3 Exploration and innovation

In Figure 5 a typical example of a punctuated equilibrium is shown. After a quick drop of the diversity at the beginning (gen. 0-3) the system enters the explorative phase, which is charactarized by high innovativity, productivity and diversity. Many new objects are generated rather continously and there are no string types with exceptionally high concentrations.

The explorative phase comes to an end, when string types appear (gen. 40) that are replicating with every other string, thus quickly dominating the soup.

The behavior can be termed a punctuated innovation which leads, in this case, to an end of the explorative phase. The system becomes simple and enters a stable phase where the organization structure is reduced only through genetic drift.

It is different from the AND reaction, where the process leading to the final string is more gradual and continuous.

## 5.4 Evolution

This section describes an experiment, that shows evolutionary behavior in the absence of external variation operators (i.e. mutation) and the absence of explicit selection[5]. The following setup has been used for the experiments in this section:

(a) Soup size $M = 10^5$ or $M = 10^6$. These large soup sizes are used to get an explorative behavior and to reduce the effects of the non-deterministic components of the reactor algorithm.

(b) Elastic collisions introduced through the filter condition

$$f_1(s_1, s_2, s_3) = (s_1 \neq s_3 \wedge s_2 \neq s_3).$$

By using this filter condition the exact replication is disabled. Every collision that normally produces a string identical to one of the reactands is considered to be elastic [**?**].

(c) Reaction mechanism: automata reaction with code table 2. Code table 2 does not contain a NOT operation. Without a NOT operation it is harder to construct a self-maintaining organization of strings under filter condition $f_1$.

(d) The soup is seeded with random strings from $\{0,1\}^{32}$, resulting in an initial diversity approximately equal to one.

Figure 6 shows the short-time behavior[6]. As in Figure 5 an explorative phase with very high diversity can be observed (generation $0 - 110$). A lot of completely new strings are produced. The productivity (fraction of collisions that are not elastic) is similar to the productivity of a random soup.

[4] A set $\mathcal{A}$ of object species is **closed** if every interaction within $\mathcal{A}$ produces only objects already in $\mathcal{A}$. A set $\mathcal{A}$ of object species is **self-maintaing** if every object is produced by at least one interaction within $\mathcal{A}$ [**?**].

[5] The selection of two objects in the reactor algorithm is not selection in the sense of evolution theory. The random selection simulates only random collisions among objects.

[6] Although this is called "short-time behavior", a rather long time is shown compared to Figure 3 .

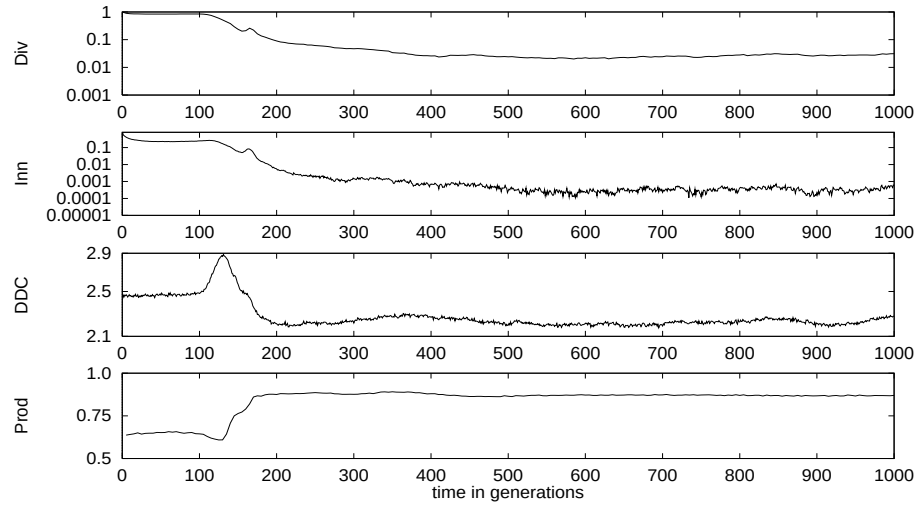



Figure 5: *Short-time evolutionary behavior. Parameters: soup size* $M = 10^5$*, automata reaction with table 2, filter condition* $f_1$*. soup seeded with* $M$ *random strings.*

This changes when a self-maintaining though not closed organization emerges (generation $110 - 200$). This does not happen as fast as in Figure 5 . Here the process takes a much longer time and seems to be more complex. Many different "species" are competing for space. Over a short period of time new strings are generated which increase their concentrations and are replaced by "better" ones. Figure 8 shows this early phase for some representative strings of a run with soup size $M = 10^6$.

Figure 7 shows the long time behavior. After the first early evolutionary phase (gen $110 - 200$) an organization has emerged whose gene diversity is massively reduced. It consists of approximately 50000 different string types where only about 3000 are present at a specific time. This organization still generates completely new strings but at a much lower frequency than the soup in the explorative phase (generation $0 - 110$).

The productivity in Figure 7 (generation $500 - 7000$) shows, that even this (over a long period stable) organization is able to develop itself. The character of this evolutionary process is different from that shown in Figure 8 . There, evolution is taking place on the bitstring (better: quasi-species) level. Here, the core set of strings is not changed. The organization is enhanced by adding/absorbing "useful" new strings which are "invented" by it.

The phenomenon that long quasi-stable periods are interrupted by rapid changes is often referred to as **punctuated equilibrium** [?]. It has been observed in natural as well as in artificial systems (e.g. DNA world [?] and multi-agent systems [?]).

For a more detailed analysis of the organization structure the reaction table can be used. Figure 9 and 10 show a part of the reaction table of the 90 most frequent strings at different stages of the evolutionary process. They demonstrate how the coupling between strings becomes closer and the amount of elastic collisions is reduced.

An interesting detail is, that a cross-over mechanism has emerged. Figure 11 shows an example where 8 "cross-overing" strings form a closed organization. The organization is so small and closed, because the cross-over is performed at a specific point.

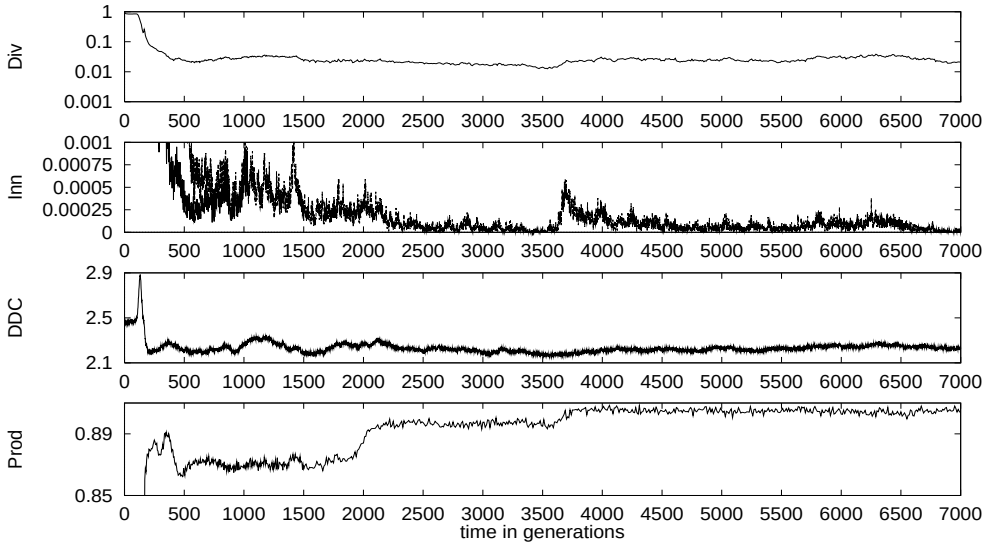



Figure 6: *Long-time evolutionary behavior. Parameters: soup size* $M = 10^5$*, automata reaction with table 2, filter condition* $f_1$*. soup seeded with* $M$ *random strings. Same run as in figure 6*

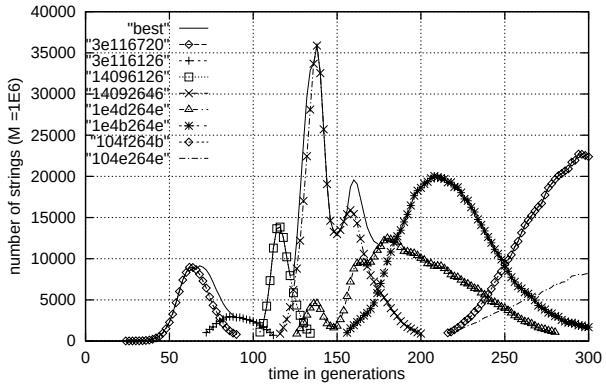



Figure 7: *Early evolutionary behavior. The concentration of some representative strings are shown. Parameters: soup size* $M = 10^6$*, automata reaction with table 2, filter condition* $f_1$*. soup seeded with* $M$ *random strings.*

generation 75

| No. | String | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 88 | 89 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8140b6fa | 29 | * | * | * | * | * | * | * | * | * |
| 1 | 140da6e8 | - | - | - | - | - | - | - | - | - | - |
| 2 | a625ffb3 | 29 | 22 | - | * | * | - | - | * | - | - |
| 3 | 140ca6f2 | - | 14 | - | - | 14 | 14 | 14 | 14 | 14 | 14 |
| 4 | 77ae526e | - | * | * | - | - | * | * | - | * | * |
| 5 | c46b12a1 | * | * | * | * | * | * | * | * | * | * |
| 6 | 4a8c05c9 | * | * | * | * | * | * | * | * | * | * |
| 7 | 223be8ae | - | - | * | - | - | * | * | - | 82 | * |
| 88 | 7ecbf629 | * | * | * | * | * | * | * | * | * | * |
| 89 | 70507121 | * | - | * | * | * | - | - | * | - | - |

generation 110

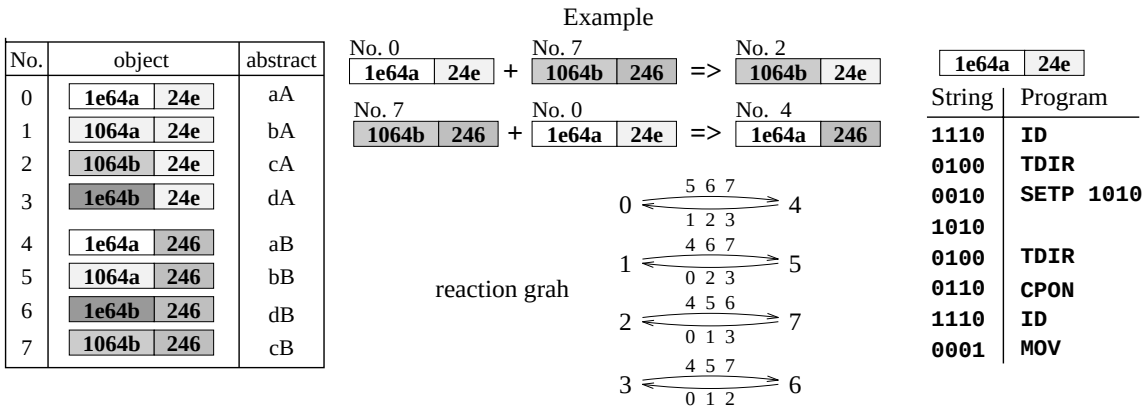
| No. | String | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 88 | 89 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | c11a260b | - | - | - | - | - | - | - | - | - | - |
| 1 | 011a2614 | 02 | - | - | - | - | 02 | 02 | 02 | 02 | 02 |
| 2 | 011a2615 | - | - | - | 01 | 01 | - | - | - | - | - |
| 3 | c11a260e | - | - | - | - | - | - | - | - | - | - |
| 4 | c11a260a | - | 00 | - | 00 | 00 | - | - | - | - | - |
| 5 | 8211e267 | 27 | 28 | 28 | 27 | 27 | - | 28 | 27 | 28 | 27 |
| 6 | 1211e267 | 23 | - | - | 23 | 23 | 23 | - | 23 | - | 23 |
| 7 | d211e267 | - | 06 | 06 | - | - | 23 | - | - | * | - |
| 88 | 211a260d | 49 | - | - | 49 | 49 | 49 | - | 49 | - | 49 |
| 89 | c611e267 | - | * | * | - | - | * | * | - | * | - |

Figure 8: *Left:Reaction table during the explorative phase (genration 75, figure 6 ). The objects are loosely coupled. Many reaction pathways leave the reaction table (denoted by "*"). Right: Reaction table shortly after the explorative phase (generation 110, figure 6 ). The objects are coupled more closely. Still many reaction pathways leave the reaction table. "-" means an ellastic collision (no reaction product).*

generation 3400

| No. | String | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 67 | 68 | 69 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 104e264b | 01 | - | 03 | 02 | 05 | 04 | 07 | 06 | 01 | - | 03 | 02 | 01 | - | 02 | - | 03 | - | 64 | 65 | - |
| 1 | 104e264a | - | 00 | 03 | 02 | 05 | 04 | 07 | 06 | - | 00 | 03 | 02 | - | 00 | 02 | 00 | 03 | 00 | 64 | 65 | 00 |
| 2 | 104f264b | 01 | 00 | 03 | - | 05 | 04 | 07 | 06 | 01 | 00 | 03 | - | 01 | 00 | - | 00 | 03 | 00 | 64 | 65 | 00 |
| 3 | 104f264a | 01 | 00 | - | 02 | 05 | 04 | 07 | 06 | 01 | 00 | - | 02 | 01 | 00 | 02 | 00 | - | 00 | 64 | 65 | 00 |
| 4 | 140e264b | - | 00 | - | 02 | - | - | - | 06 | 00 | 00 | 02 | 02 | 00 | 00 | 02 | 00 | 02 | 00 | - | 65 | 00 |
| 5 | 140e264a | 01 | - | 03 | - | - | - | 07 | - | 01 | 01 | 03 | 03 | 01 | 01 | 03 | 01 | 03 | 01 | 64 | - | 01 |
| 6 | 140f264b | - | 00 | - | 02 | - | 04 | - | - | 00 | 00 | 02 | 02 | 00 | 00 | 02 | 00 | 02 | 00 | - | 65 | 00 |
| 7 | 140f264a | 01 | - | 03 | - | 05 | - | - | - | 01 | 01 | 03 | 03 | 01 | 01 | 03 | 01 | 03 | 01 | 64 | - | 01 |
| 8 | 104e246b | 09 | - | 11 | 10 | 30 | 24 | 31 | 28 | 09 | - | 11 | 10 | 09 | - | 10 | - | 11 | - | * | * | - |
| 9 | 104e246a | - | 08 | 11 | 10 | 30 | 24 | 31 | 28 | - | 08 | 11 | 10 | - | 08 | 10 | 08 | 11 | 08 | * | * | 08 |
| 10 | 104f246b | 09 | 08 | 11 | - | 30 | 24 | 31 | 28 | 09 | 08 | 11 | - | 09 | 08 | - | 08 | 11 | 08 | * | * | 08 |
| 11 | 104f246a | 09 | 08 | - | 10 | 30 | 24 | 31 | 28 | 09 | 08 | - | 10 | 09 | 08 | 10 | 08 | - | 08 | * | * | 08 |
| 12 | 104e206b | - | 00 | - | 02 | 00 | 00 | 02 | 02 | - | 08 | - | 10 | - | - | 16 | 21 | - | 18 | 00 | 02 | 25 |
| 13 | 104e206a | 01 | - | 03 | - | 01 | 01 | 03 | 03 | 09 | - | 11 | - | - | - | - | - | 14 | - | 01 | 03 | 23 |
| 14 | 104f206a | 01 | - | 03 | - | 01 | 01 | 03 | 03 | 09 | - | 11 | - | 13 | - | - | - | - | - | 01 | 03 | 23 |
| 15 | 104e261a | 01 | - | 03 | - | 01 | 01 | 03 | 03 | 09 | - | 11 | - | 13 | - | - | - | 14 | - | 01 | 03 | 23 |
| 16 | 104f206b | - | 00 | - | 02 | 00 | 00 | 02 | 02 | - | 08 | - | 10 | - | 12 | - | 21 | - | 18 | 00 | 02 | 25 |
| 17 | 104e260a | 01 | - | 03 | - | 01 | 01 | 03 | 03 | 09 | - | 11 | - | 13 | - | - | - | 14 | - | 01 | 03 | 23 |
| 67 | 1046264b | 01 | 00 | 03 | 02 | 05 | 04 | 07 | 06 | * | * | * | * | * | * | * | 00 | * | 00 | 64 | 65 | 00 |
| 68 | 10af264a | 59 | 59 | - | - | 59 | 59 | - | - | * | * | * | * | * | * | * | * | * | * | 59 | - | * |
| 69 | 104e265e | 01 | - | 03 | - | 01 | 01 | 03 | 03 | 09 | - | 11 | - | 13 | - | - | - | 14 | - | 01 | 03 | - |

Figure 9: *Reaction table of a converged soup (generation 3400, Figure 6 ). The objects are coupled very closely. Only a few reaction pathways are leaving the reaction table. The objects are very similar.*




| No. | object | abstract |
|---|---|---|
| 0 | 1e64a 24e | aA |
| 1 | 1064a 24e | bA |
| 2 | 1064b 24e | cA |
| 3 | 1e64b 24e | dA |
| 4 | 1e64a 246 | aB |
| 5 | 1064a 246 | bB |
| 6 | 1e64b 246 | dB |
| 7 | 1064b 246 | cB |

Figure 10: *Emergence of cross-over. The strings are able to insert a part of their sequence into the sequence of an arbitrary string. The result is a cross-over of both. E.g.* $1e64a24e + 00000000 \Longrightarrow 0000024e$.

# 6 Summary and Discussion

- We have extended the works mentioned above by introducing a reaction mechanism inspired by techniques of computer science. For exploring generalizations about evolving systems, the high speed at which the reactions are carried out is helpful, because it allows interactive and large experiments. the disadvantages are the (so far) fixed length of strings and that the formalization of the reaction mechanism is more complicated than using lambda-calculus [**?**] or matrix multiplication [**?**].

- It has been demonstrated, that the building blocks used in computer science to describe hardware elements and functions can be easily used to construct self-organizing and even evolving systems.

- Macroscopic monitoring techniques have been used. They are able to show qualitative changes of the system's dynamics. But for answering detailed questions they are far from sufficient. Design of visualisation techniques, expecially interactive ones, is needed for future research.

- The experiments show that complex forms of evolution can take place in systems without any explicit variation operator, like mutation or recombination, and without any explicit (artificial) selection.

- The setup used here is one of the simplest ways of letting machines acting on each other. We think that an important concept concerning natural evolution is, that the way evolution is carried out is itself under the influence of evolution, a concept which is called **meta-evolution**. Meta-evolution has long been applied in evolutionary algorithms. For example, in evolution strategies (ES) parameters are used which determine the variance and covariance of a generalized n-dimensional normal distribution for mutations. The strategy parameters themselves are adapted during the optimization process. But the operators used for adaptation are fixed.

  Here, a variation of an object alters the way how other objects are modified, and so on. We can interpret such a system, as having many levels of meta-evolution, maybe a potentially

infinite number of levels.

- The complex behavior aquired here, results from a simple, ad hoc hand-written interaction mechanism working on small objects in an environment with no spatial structure. Introducing larger, variable-length objects and topological strucures, allowing the formation of niches, will result in an increase in complexity. Then we have to face the danger, that in order to grasp a world we do not understand, we create a system we do not understand [?, ?]. To circumvent this, powerfull analyzation and visualization methods are required, that should be general (like the DDC), so that they can be applied to different systems.
- Reaction systems – like the system discussed here – are able to perform computations [?]. For example a chemical reaction system has been used to instantiate an artificial neural network [?] or to solve a variant of the Hammilton-path problem [?]. So far, the reaction pathways have been chosen carefully by hand. The integration of self-evolution into molecular computing systems is promising and will be an aim of future research.

## Acknowledgements

This project is supported by the DFG (Deutsche Forschungsgemeinschaft), grant Ba 1042/2-1. We also thank Helge Baier, Peter Nordin and Jochen Triesch for many stimulating discussions and Ulrich Hermes for his outstanding technical support.

# APPENDIX

The Appendix contains a description for all registers and instructions of the automata reaction.

### 6.0.1 The control registers

In addition to the two 32 bit registers the automaton has 5 smaller **control registers**. The control registers are initialized with the first value given in the following description:

direction $r_D$ — $r_D \in \{\text{left}, \text{right}\}$
The direction register determines the direction the pointers are moving. The operation TDIR toggles the direction.

move-mode $r_M$ — $r_M \in \{\text{both}, \text{IOpointer}, \text{operator-pointer}\}$
This register controls which pointer is moved in case of a pointer movement. The operation TMM toggles the movement mode.

last-ALU-operation $r_{ALU}$ — $r_{ALU} \in \{ID, NOT, AND, OR, EXOR, EQ\}$
This register stores the last ALU operation. If the copy mode is switched on, the operation stored in $r_{ALU}$ will be executed before each movement step.

copy-mode $r_C$ — $r_C \in \{\text{off}, \text{on}\}$
If the copy mode is switched on, the last ALU operation stored in

$r_{ALU}$ is executed before each movement step of the IO pointer and the operator-pointer.

**pattern** $r_P$ — $r_P \in \{\text{none}\} \cup \{0,1\}^4$
With "SETP **p**" this register will be set to the pattern **p** (**p** $\in \{0,1\}^4$). The operation UNSETP clears the pattern by setting $r_P$ to none. If the pattern is set, each MOV operation will move the pointers unless the pattern is found in the porgramm register (max. 32 steps). If only the IO pointer is moved the pattern is searched in the IO register. If $r_P =$ none the pointers are moved one step.

### 6.0.2 Arithmetic and logic instructions

During an arithmetic or logic operation bit $b$ and $b'$ are processed by the ALU (Figure 2 ). The result overwrites $b'$. Then the pointers are moved accordingly to the direction register $r_D$ (left or right) and the the move-mode register $r_M$ (both, IO pointer only or operator pointer only).

**ID** identity: $b' := b$, $r_{ALU} := ID$.

**NOT** negation: $b' := \neg b$, $r_{ALU} := NOT$.

**AND** logic and: $b' := b \wedge b'$, $r_{ALU} := AND$.

**OR** logic or: $b' := b \vee b'$, $r_{ALU} := OR$.

**EXOR** execlusive or: $b' := b \neq b'$, $r_{ALU} := EXOR$.

**EQ** equality: $b' := b = b'$, $r_{ALU} := EQ$.

After the execution of a logic instruction, register $r_{ALU}$ is set to the operation executed. It is used only during the execution of the MOV instruction, if the copy mode is switched on ($r_C = on$).

### 6.0.3 Movement and control instructions

**MOV** Moves the pointers accordingly to the move-mode, direction and pattern register. If no pattern is set, ($r_P =$ none) the pointers are moved one step (one bit). If a pattern is set, the pointers are moved unless the pattern is found (max. 32 steps). If $r_M =$ IOpointer the pattern is searched in the IO register, otherwise in the program-register. Again, the direction of the movement is given by the direction register $r_D$.
If the copy-mode is switched on ($r_C :=$ on) by CPON, the last ALU operation is executed before each step.

**CPON** Switches the copy-mode on ($r_C :=$ on).

**CPOFF** Switches the copy-mode off ($r_C :=$ off).

SETP **p** Sets the pattern $\mathbf{p} \in \{0,1\}^4$ for the move operation ($r_P := \mathbf{p}$).
This is the only operation with an argument.

UNSETP Clears the pattern ($r_P :=$ none).

TDIR Toggles the move direction for the pointers.

$$r_D := \begin{cases} \text{right} & \text{if } r_D = \text{left}, \\ \text{left} & \text{if } r_D = \text{right} \end{cases}$$

TMM Toggles the move-mode:

$$r_M := \begin{cases} \text{both} & \text{if } r_M = \text{operator-pointer}\ , \\ \text{IOpointer} & \text{if } r_M = \text{both}, \\ \text{operator-pointer} & \text{if } r_M = \text{i}\text{ø}\text{-pointer} \end{cases}$$

NOP No operation.

STOP Stops the automaton before the whole program is executed.

For a precise formal specification of the automata reaction the source code is available [**?**].