Modelling and Analysing the Information Processing Capabilities of Simple Biological Systems

M. Moškon and M. Mraz

Modelling and analysing the information processing capabilities of simple biological systems

Miha Moškon and Miha Mraz

Faculty of Computer and Information Science, University of Ljubljana
 Tržaška cesta 25, SI-1000, Slovenia
 E-mail(*corresp.*): `miha.moskon@fri.uni-lj.si`
 E-mail: `miha.mraz@fri.uni-lj.si`

Abstract. Biological systems that present basic logic primitives for information processing have already been realized. Models for simulating their dynamics have also been implemented. However there is a lack of metrics that would objectively evaluate the information processing capabilities of these primitives and possibilities of their interconnectivity. With the introduction of such processing and performance descriptive quantities complex biological systems capable of information processing could be built more straightforwardly. That would bring us closer to the realization of a biological computer.

Keywords: models for synthetic biology, stochastic modelling, chemical master equation, metrics, unconventional computing.

AMS Subject Classification: 93A30; 68U20.

1 Introduction

The idea of processing in biological systems has proved its potential for the first time two decades ago with the field known as *DNA computing* [1]. On the other hand *synthetic biology* is a rapidly evolving field that incorporates different methods for synthesis of new biological systems that reflect the behaviour in accordance with defined functionalities. These systems can also perform as various information processing blocks such as *logic gates* [19], *oscillators* [5] or *flip-flops*. Biological computers could therefore be constructed from such primitives. In order to design more complex structures, different models that include the biochemical properties and reflect the dynamics of such biological systems can be used. The main weakness in their usage is in our opinion in the absence of metrics that would estimate the suitability of their behaviour as an information processing platforms based on their dynamics. We present the introduction of such metrics and the demonstration of their evaluation on an

example of stochastic model of a simple biological system, i.e. toggle switch [7]. Analysis of its dynamics from the information processing perspective on the basis of evaluated metrics is also demonstrated. Toggle switch circuit is on the one hand complex enough to demonstrate the metrics evaluation process. On the other hand it has already been studied extensively and many data regarding its properties can be found in the literature.

The rest of the paper is organized as follows. Section 2 explains the dynamics in modern computer systems. In section 3, we introduce the field of information processing in biological systems and present the model of RS flip-flop biological equivalent, i.e. toggle switch. In section 4 modelling of such systems is described. Section 5 presents the establishment of metrics for analysing these systems from an information processing viewpoint. Evaluation of metrics established is demonstrated on an example of presented toggle switch model. Section 6 is devoted to analysis of an example model which is performed based on metrics evaluated. Finally, section 7 concludes the paper.

2 Dynamics in modern computer systems

Modern computer systems and their components are mainly based on binary logic. Signal values in such systems can be presented by two logical levels, namely low level (logical 0) and high level (logical 1). Such interpretation can be achieved with the binarization of a physical quantity that encodes the information (e.g. voltage in semiconductor circuits). The behaviour of such systems can be presented by a sequence of transitions among logical level 0 and logical level 1. These transitions can be caused intentionally with some external influence (*logical switching*) or can be triggered unintentionally by so-called *noise*. After the transition from one logical level to another, state of the system can also be *memorized* for some longer period.

2.1 Logical switching

Switches among different states of the system are usually caused by some external influence, i.e. by the change of some input signal. While we are dealing with a binary system, only two types of transitions are possible. Ideally these switches would be performed without any delay, but in reality the *transition times* must be regarded. Time needed to perform a switch from low level to high level is so-called *rise time* and time to perform a switch from high level to low level is so-called *fall time* [17, 18]. Rise and fall times are usually different. When logical switch is being performed the binary interpretation of observed signal can not be performed while the signal is in the region that belongs neither to logical level 0, neither to logical level l. This region is so-called *region of uncertainity* (see Section 2.3). Signal value has a valid interpretation after the time, which is bigger than maximal time needed to stabilize the signal, i.e. *switching time*.

2.2 Memorizing the logical state

Memorization of the logical state is a capability of so-called *memory circuits*. Memory circuits can be roughly divided in two groups: memory circuits that are able to sustain their state for unlimited amount of time (e.g. SRAM circuits) and memory circuits that need *refreshing* with some output signal in order to sustain their state (e.g. DRAM circuits).

In order to realize memory circuits in binary logic, so-called *bistable circuits* are used. Bistable circuits have two stable states, where each stable state represents different logical level. State change is performed by input signal. During the switch from one state to another the circuit can stuck in a state between stable states and does not have a valid interpretation. On the other hand transition to this state can be also caused if refresh rate of memory circuits that need refreshing is to low.

2.3 Dynamic caused by noise

Deviation of signal from its ideal value is so-called *noise* and can be caused by fluctuations within the system (*intrinsic noise*), or by external influences on the system (*extrinsic noise*) [18]. Noise is absent in ideal signal, but in reality noise is always present. It can be removed only in the limited scope. On the other hand its influences on the interpretation of the signal can mostly be eliminated.

The main problem is that noise can cause misinterpretations of the logical state of the system. The direct result is loss of the information and thus transition from deterministic to indeterministic system. Our system therefore becomes useless from an information processing point of view. We can solve this problem with proper interpretation of the physical quantity that encodes the information. Interpretation is appropriate if maximal noise does not cause the transition of the system from one logical state to another. This can be achieved with so-called *noise margin* introduction, which defines the region in which the signal is interpreted as a logical level 0 and the region in which the signal is interpreted as a logical level 1. Between these regions signal does not have a valid interpretation, therefore this region is so-called *region of uncertainity*. Regions must be defined appropriately in order to trigger the transitions among them only by the intentional change of input signal. Influence of input signals must be higher than the influence of maximal possible noise.

3 Realization of RS flip-flop biological equivalent

### 3.1 Dynamics in biological systems

Presence of specific *protein* can be interpreted as an information from the information processing point of view. It can be quantitatively presented with its *concentration* which is for our purposes measured in nanomoles per litre (nM). Protein concentration dynamics is in the type of the systems we analyse, i.e. *gene regulatory networks*, a consequence of two processes, i.e. its *synthesis* and its *degradation*.

Protein synthesis can be described with two processes, i.e. *transcription* and *translation*. In the process of transcription DNA is transcribed to *messenger RNA* (mRNA). Protein is synthesized in translation process on the basis of mRNA sequence. Rate of transcription is controlled by so-called *transcription factors* which bind to specific DNA sequences, i.e. *promoters*. Binding of transcription factors to promoter can activate/increase or repress/decrease the transcription rate. Transcription factors that increase the transcription rate are so-called *activators* and transcription factors that decrease the transcription rate are so-called *repressors*. We can presume that translation is a process that is performed unconditionally. Protein synthesis is therefore strictly dependent on the absence, respectively presence, of transcriptional factors. Protein synthesised can also perform as a transcriptional factor for the synthesis of some other protein or even for its own synthesis. With the modifications in promoter sequence of DNA we can define which proteins will act as transcriptional factors and how will they affect the transcription (i.e. activate or repress) of the protein encoded in the rest of the DNA sequence. Systems constructed of several DNA sequences which interact among each other can thus be constructed. The network of such DNA sequences is called *gene regulatory network*. An example of such gene regulatory network is a network of two mutually repressing proteins, where first protein represses the synthesis of the second protein and vice versa.

Degradation causes the reduction of protein concentration and is therefore a process opposite to synthesis. Its rate can be controlled with the introduction of external influences (e.g. UV light) but it is always present in certain amount.

3.2   *RS* flip-flop

RS *flip-flop* is one of the simplest possible memory elements in semiconductor technology [11]. Its circuit is composed of two NOR (NOT OR) gates, where the output of each one is an input to the other one. Logical scheme of the circuit is presented in Figure 1(a).

R	S	D^1q	$D^1\overline{q}$
0	0	q	$\overline{q}$
0	1	1	0
1	0	0	1
1	1	?	?

(a)   (b)
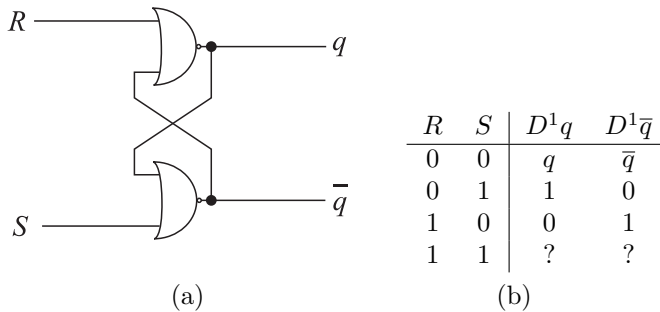
Figure 1. Logical scheme of RS flip-flop is presented in Figure (a) and its behaviour in Figure (b), where R presents Reset input, S Set input, q and $\overline{q}$ current outputs of the flip-flop and D^1q in $D^1\overline{q}$ outputs of the flip-flop in the next time step. Undefined output value is denoted with ?.

The outputs of NOR gates also present the outputs of the RS flip-flop circuit. Therefore there are two outputs where the second one ($\overline{q}$) is the negation of the first one (q). We can control the circuit's behaviour with two inputs, i.e. *Reset* input (R) which deactivates output q (sets its value to logical 0) when active and *Set* input (S) which activates output q (sets its value to logical 1) when active. If both inputs are absent, current state of the system does not change - state of the system is memorized. Presence of both inputs is forbidden and causes undefined behaviour of the circuit. Circuit's behaviour at different input combinations is presented in the logical table in Figure 1(b).
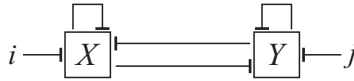
3.3 Toggle switch

Analysis of switching dynamics will be performed on the model of RS flip-flop biological equivalent, i.e. *toggle switch* [3, 7]. Toggle switch circuit is complex enough to demonstrate the metrics evaluation process which is described in section 6. On the other it has already been studied extensively. Therefore many data regarding its behaviour already exist in the literature.

Toggle switches exist in many variations. Our analysis will be based on a toggle switch which originates in the switching mechanism of the Bacteriophage λ [3]. We can present its structure as a gene regulatory network of two mutually repressing proteins. In order to limit uncontrolled growth of their concentrations, each of them also represses itself (negative feedback), but only when its concentration is very high.

Stable state of the system is presented with high concentration of exactly one observed protein. While there are two possible stable states we are dealing with a bistable system, which can be used as a memory element. Switches among stable states can be triggered with an external input (e.g. UV light) similar as with RS flip-flop. In our example external input increases the degradation rate of one of the observed proteins [3]. Concentration of the protein with increased degradation rate therefore drastically decreases. Repression of the other protein is stopped and rate of its synthesis is increased. State of the system which is defined with high concentration of exactly one observed protein is therefore switched if degradation rate of protein with high presence is increased or refreshed if degradation rate of protein with low presence is increased. We can presume that we have two different external influences (two different light spectres) at our disposal where each of them affects only one of the observed proteins.

Biological scheme of described system is presented in Figure 2, where X and Y present proteins observed and i and j present the presence of external inputs which affect the degradation rate of each of the observed proteins.

If we presume that low protein concentrations present logical value 0 and high concentrations logical value 1, presence of proteins X and Y can be interpreted as outputs of the system, analogous to signals q and $\overline{q}$ in RS flip-flop. Respectively external influences that increase certain protein degradation rates, i.e. i and j, can be interpreted as Reset (R) and Set (S) inputs. Toggle switch dynamics is therefore in accordance with RS flip-flop logical behaviour.

Figure 2. Biological scheme of gene regulatory network presenting the toggle switch circuit, where X and Y present proteins observed and i and j present external influences on the behaviour of the circuit. Connections among different parts of the circuit present their mutual repression.

4   Modelling the dynamics of biological systems

When the number of different chemical species and the number of different reactions in biological system is increased its complexity drastically grows. Different models can be used in order to be able to predict their behaviour and therefore decrease the complexity and amount of experimental work [2, 9]. Models can also be used for prediction of the switching dynamics within the observed biological system and thus analyse its suitability as an information processing element. In this way basic properties of switching dynamics within the system can be obtained without experimental work.

4.1   Modelling basics

Models used for modelling the dynamics of biological systems can be roughly divided to *deterministic* [5, 7], *stochastic* [6, 10, 12, 15] and *semiquantitative* [9]. Deterministic models are relatively simple, but can only be used in order to approximate the average system response, or individual system response measured in a long time period. Modelling the heterogeneity of the population of identical biological systems can be performed using stochastic models which incorporate stochasticity of biochemical processes and influence of noise on the system. Mathematical complexity of solving the stochastic models is much higher than the complexity of deterministic models. In order to decrease it and still have an approximation of population heterogeneity, semiquantitative models can be used as a compromise. Different types of models exist within these groups, but deterministic models are mostly based on ordinary differential equations (ODEs), i.e. Hill equations [2] and stochastic models are mostly based on solving the Chemical Master Equation (CME) [4]. While we are dealing with relatively simple biological system, stochastic model will be used here in order to predict its switching dynamics.

4.2   Modelling the dynamics of a toggle switch example

We can present the state of the system as a vector of non-negative real numbers describing concentrations of chemical species observed [4]. We presume that chemical species presented in Table 1 [3] define the state of the analysed system. Preferable dynamics of the system is achieved if the final form of observed proteins (i.e. X and Y) is their *dimerized* form (i.e. X_2 and Y_2). Proteins therefore affect the process of transcription after the dimerization. Model described here is mostly based on the model presented in [3].

chemical specie	role in the system
DNA_X	promoter of DNA encoding protein X
DNA_Y	promoter of DNA encoding protein Y
X	protein X
Y	protein Y
X_2	dimer of protein X
Y_2	dimer of protein Y
X_2DNA_X	DNA_X bounded with protein dimer X_2
Y_2DNA_Y	DNA_Y bounded with protein dimer Y_2
Y_2DNA_X	DNA_X bounded with protein dimer Y_2
X_2DNA_Y	DNA_Y bounded with protein dimer X_2

Table 1. Chemical species that are observed within the analysed system.

Change of the vector describing concentrations of observed chemical species is performed in discreet time steps in dependence on observed chemical reactions within the system. Certain chemical reaction could be triggered within each time interval of the simulation and therefore change the state of the system regarding the triggered reaction's properties. Table 2 presents a list of observed chemical reactions.

Reaction rates were evaluated regarding the relevant literature [3, 13, 16]. Preferable dynamics of the system (symmetrical switching, memorizing the state, concentration levels) is achieved when reaction rates for both observed proteins (i.e. X and Y) are adapted from protein CI in *Bacteriophage* λ [3].

Stochastic model, i.e. Chemical Master Equation, was defined on the basis of defined chemical species, chemical reactions and their rates.

4.3 Solving the Chemical Master Equation

Chemical Master Equation (CME) defines the probability that the state of the system in certain time, i.e. $X(t)$, will equal state x, given that initial state of the system $X(t_0)$ equals state x_0

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t} = \sum_{j=1}^{M} \left[a_j(x - \nu_j) P(x - \nu_j, t|x_0,t_0) - a_j(x)P(x,t|x_0,t_0) \right]$$

(4.1)

where x and x_0 present vectors of non-negative real numbers describing concentrations of observed chemical species, M number of reactions observed, $a_j(x)$ probability of reaction j to occur at the state x, and ν_j state change vector, which determines the modification of vector x when reaction j occurs.

CME therefore presents a system of differential equations. Its size increases proportional to the number of combinations of chemical species within the system. While it can be solved only for very small systems [4], different approximative approaches are used. In order to solve it, we used Gillespie algorithm for stochastic simulation, i.e. Stochastic Simulation Algorithm (SSA) [8].

SSA calculates the time evolution of the state of the system, i.e. $X(t)$, regarding the following equation, which defines the probability of reaction j in

reaction	description
$DNA_X \rightarrow X + DNA_X$	synthesis of protein X
$DNA_Y \rightarrow Y + DNA_Y$	synthesis of protein Y
$X + X \rightarrow X_2$	dimerization of protein X
$Y + Y \rightarrow Y_2$	dimerization of protein Y
$X_2 \rightarrow X + X$	monomerization of dimer X_2
$Y_2 \rightarrow Y + Y$	monomerization of dimer Y_2
$X_2 + DNA_X \rightarrow X_2DNA_X$	self-repression of X
$Y_2 + DNA_Y \rightarrow Y_2DNA_Y$	self-repression of Y
$Y_2 + DNA_X \rightarrow Y_2DNA_X$	cross-repression of X
$X_2 + DNA_Y \rightarrow X_2DNA_Y$	cross-repression of Y
$X_2DNA_X \rightarrow X_2 + DNA_X$	dissociation of X and DNA_X
$Y_2DNA_Y \rightarrow Y_2 + DNA_Y$	dissociation of Y and DNA_Y
$Y_2DNA_X \rightarrow Y_2 + DNA_X$	dissociation of Y and DNA_X
$X_2DNA_Y \rightarrow X_2 + DNA_Y$	dissociation of X and DNA_Y
$X \rightarrow \emptyset$	degradation of protein X
$Y \rightarrow \emptyset$	degradation of protein Y
$X_2 \rightarrow \emptyset$	degradation of dimer X_2
$Y_2 \rightarrow \emptyset$	degradation of dimer Y_2
$X_2DNA_X \rightarrow DNA_X$	degradation of bounded dimer X_2
$X_2DNA_Y \rightarrow DNA_Y$	degradation of bounded dimer X_2
$Y_2DNA_X \rightarrow DNA_X$	degradation of bounded dimer Y_2
$Y_2DNA_Y \rightarrow DNA_Y$	degradation of bounded dimer Y_2

Table 2. Observed chemical reactions that directly affect the dynamics of the analysed system [3].

time interval $[t, t + \tau]$.

$$p(\tau, j | x, t) = a_j(x)e^{-a_0(x)\tau} \tag{4.2}$$

Equation 4.2 implies that the time step τ and the reaction j can be selected according to expressions

$$\tau = \frac{1}{a_0(x)} \ln \left(\frac{1}{r_1} \right) \tag{4.3}$$

$$\sum_{j'=1}^{j-1} a_{j'}(x) \leq r_2 a_0(x) < \sum_{j'=1}^{j} a_{j'}(x) \tag{4.4}$$

where r_1 and r_2 are random numbers from the uniform distribution in the unit-interval [4].

Following steps are performed in order to calculate the time evolution of the state of the system:

1. *Initialization*: initialize the time and the state of the system ($x = x_0$, $t = t_0$).

2. *Evaluate the reaction probabilities*: regarding the current state x and reaction rates, calculate the probability that reaction j occurs in the next time interval ($a_j(x)$). Also calculate the sum of all reaction probabilities ($a_0(x)$).

3. *Monte Carlo step*: generate values r_1 and r_2 and determine the time step τ and reaction j according to Equations 4.3 and 4.4.

4. *State update*: according to step 3 update the time $(t \leftarrow t + \tau)$ and state of the system $(x \leftarrow x + \nu_j)$.

5. *Iteration*: return to step 2 or end the simulation if t has exceeded defined simulation time.

5   Introduction of metrics for evaluating the dynamics of biological systems

In order to objectively evaluate the dynamics of biological system for the purposes of information processing platforms, different metrics were introduced. Their values were evaluated based on the simulation results of an example of toggle switch model described in Section 3.3. Metrics were established on the basis of metrics used in classical, i.e. semiconductor digital circuits [17, 18]. Based on the results of the simulations following metrics were evaluated:

- determination of chemical specie that encodes the information,

- evaluation of noise,

- evaluation of refresh rate,

- evaluation of logical levels,

- evaluation of switching time.

Evaluation of their values on the basis of simulation results of toggle switch model will be presented in the following section. Evaluated values are highly dependent on the properties of the system used, i.e. on the observed chemical species, reactions among them and kinetic rates of the reactions. Numerical values are therefore strictly bound to chemical species and reactions observed in modelled system. On the other hand principles of their evaluation are independent on the properties of analysed biological system.

5.1   Determining the chemical specie that encodes the information

Chemical specie and appropriate physical quantity that measures its presence in the system have to be determined in order to encode the information within the biological system. Chemical specie chosen for the role of encoding the information should perfectly reflect desired dynamics of the system and should also be measurable with appropriate physical quantity. Chosen physical quantity should quantitatively reflect the presence of chemical specie and should have a similar role as voltage in semiconductor circuits. It turns out that it is appropriate to choose physical quantity *concentration* for this purpose while its properties are in accordance with our demands.

As already mentioned concentration has to be bound to a certain set of chemical species where the dynamics of the species chosen should be in accordance with the dynamics of RS flip-flop described in section 3.2. Their desired properties are also maximal *signal to noise ratio* and minimal *switching time* [17, 18].

Regarding the observed chemical species in our system potential candidates that could encode the information are *protein monomers* (i.e. X and Y), *protein dimers* (i.e. X_2 and Y_2) and *protein bounded to DNA of complementary protein* (i.e. X_2DNA_Y and Y_2DNA_X). Time evolution of concentrations of potential candidates in the simulation of switching dynamics of the biological system is presented in Figure 3, where Figure 3(a) presents protein monomers, Figure 3(b) protein dimers and Figure 3(c) pair of proteins bounded to DNA of complementary protein.

(a)

(b)

(c)
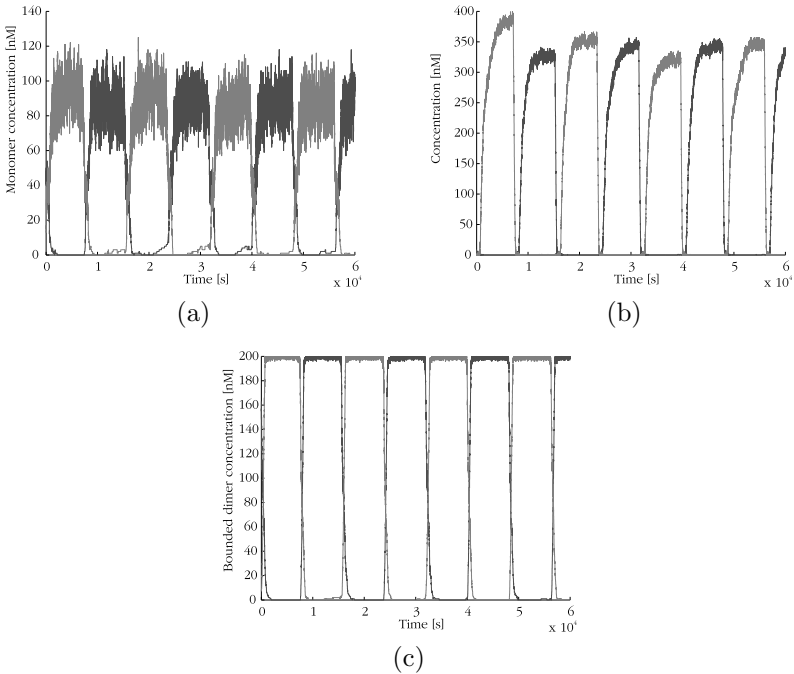
Figure 3. Time evolution of concentration of three groups of chemical species which present potential candidates that could encode the information in RS flip-flop biological equivalent. Figure (a) presents the concentration of monomers, Figure (b) concentrations of dimers and Figure (c) concentrations of pair of proteins bounded to DNA of complementary protein. Concentrations of the chemical species that are directly associated with protein X are presented with dark-grey line and concentrations of the chemical species that are directly associated with protein Y with light-grey line.

Simulation results indicate that the most appropriate chemical specie that could encode the information is pair of proteins bounded to DNA of complementary protein, i.e. X_2DNA_Y and Y_2DNA_X, Figure 3(c). The problem lies in the fact that these two chemical species can not be used as an output of

the system, while they only present its internal state. On the other hand protein monomers and dimers could be used as transcriptional factors to some other gene regulatory network and therefore present valid outputs of the system. While the dimer properties are much better regarding the earlier defined demands, X_2 and Y_2 are chosen in order to encode the information.

Time evolution of protein dimers concentrations, i.e. concentrations of X_2 and Y_2, will be observed in dependence on the presence of external influences, i.e. i and j. Simulated system dynamics can be compared with the dynamics of RS flip-flop, where protein dimers X_2 and Y_2 will have a role of RS flip-flop outputs, i.e. q and $\bar{q}$, and external influences i and j will have a role of RS flip-flop inputs, i.e. R and S.

5.2 Evaluating the noise

We are only considering intrinsic noise in our model which arises from the fluctuations inherent to our system, i.e. from fluctuations in transcription and translation [10]. The aim of the noise values evaluation is determination of *noise margins* and *region of uncertainity*. It turns out that noise is maximal when the system is memorizing its internal state. Different noise values were measured for the high level signal (i.e. logical 1) and different values for the low level signal (i.e. logical 0). Noise values were measured for both signals encoding the information (i.e. X_2 and Y_2) and were approximately the same for both of them. It was estimated that maximal noise present in high level signal equals 100 nM and maximal noise present in low level signal 50 nM. Both values present the deviations from the average dimer concentrations and are specific for our parameter set. In order to avoid the misinterpretations of the logical values which can cause loss of the information, region of uncertainity should be at least 150 nM wide.

5.3 Evaluating the refresh rate

Stable state (exactly one of the observed proteins has high concentration) is sustained for some time after its establishment which is followed by the temporal presence of external influence (i.e. signals i or j). Eventually concentration of protein dimer with high presence ($C_H >> 0$) in the system starts decreasing and concentration of protein dimer with low presence ($C_L \approx 0$) starts increasing. State of the system starts transitioning to the state that does not have a valid logical interpretation, i.e. concentrations of both protein dimers are approximately the same (Figure 4(a)).

Figure 4(a) indicates that it is necessary to introduce refreshing in order to effectively memorize the state of the system, which is similar to dynamic memory circuits (such as DRAM circuits). Without refreshing system transitions to the state that does not have a valid logical interpretation. We presume that refreshing needs to be performed when the distance between dimer concentrations becomes smaller than the width of the uncertainty region (i.e. $C_H - C_L < 150nM$). Refresh rate (t_{REF}) can be evaluated on the basis of simulation results of our toggle switch model to $1.4 \cdot 10^5$ seconds. We can evaluate the refresh frequency (f_{REF}) as t_{REF}^{-1} to $7.15 \cdot 10^{-6} Hz = 7.15 \mu Hz$.
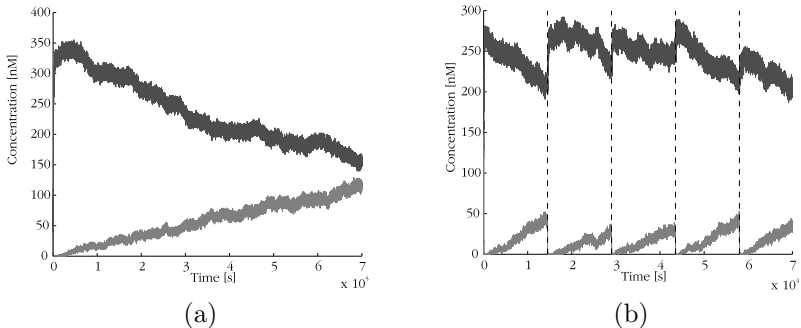
Figure 4. Time evolution of concentrations of proteins in dimerized form, where initially system is in its stable state and simulation is ran for $7 \cdot 10^5$ seconds. Figure (a) presents scenario without refreshing and Figure (b) with refreshing by temporal presence of external signals i, respectively j. Concentrations of protein dimer X_2 are presented with dark-grey line and concentrations of protein dimer Y_2 with light-grey line.

Simulation results of effective state memorization with upper defined refresh rate are presented in Figure 4(b). Details of refreshing the state of the system with temporal introduction of external influence are presented in Section 6.2.

5.4 Evaluating the logical levels

We presume that high protein dimer concentrations (C_H) can be interpreted as logical level 1 (high level), and low dimer concentrations (C_L) as a logical level 0 (low level). Logical levels can be defined accurately in the same way as logical levels are evaluated in semiconductor circuits [17, 18]. Following boundaries have to be determined:

- *high level input concentration*: minimum concentration level required for logical value 1 at an input ($C_{IH}(\min)$),

- *low level input concentration*: maximum concentration required for logical value 0 at an input ($C_{IL}(\max)$),

- *high level output concentration*: minimum concentration for logical value 1 at an output ($C_{OH}(\min)$),

- *low level output concentration*: maximum concentration for logical value 0 at an output ($C_{OL}(\max)$).

Protein dimer concentration has a valid interpretation if its values are within the regions $[0, C_{OL}(\max)]$ or $[C_{OH}(\min), \infty)$. Region of uncertainty can therefore be defined as an interval $(C_{OL}(\max), C_{OH}(\min))$. Example of time evolution of protein dimer concentrations simulation with marked boundaries that have to be determined is presented in Figure 5.
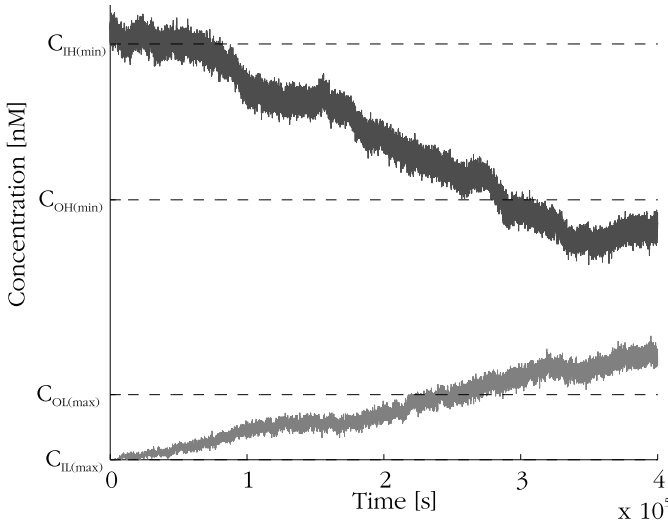
Figure 5. Example of time evolution of protein dimer concentrations simulation with denoted boundaries which define the regions for signal interpretation.

5.5 Evaluating the levels of input concentrations

Minimum protein dimer concentration level required for logical value 1 at an input ($C_{IH}(\mathrm{min})$) presents minimal concentration that must be reached after the switch or refresh of the state is performed by the protein which should be interpreted as a logical value 1. Respectively maximum concentration required for logical value 0 at an input ($C_{IL}(\mathrm{max})$) presents maximal concentration reached by the protein interpreted as a logical value 0 after the switch or refresh of the state. Both values can be evaluated with the measurements of concentration levels after the stable state is achieved due to the switch or due to the refresh of the system's state. Based on the results of the simulations $C_{IH(\mathrm{min})}$ was estimated to 320 nM and $C_{IL}(\mathrm{max})$ to 5 nM.

5.6 Evaluating the levels of output concentrations

Minimum concentration for logical value 1 ($C_{OH}(\mathrm{min})$) and maximum concentration for logical value 0 at an output ($C_{OL}(\mathrm{max})$) are boundary concentrations that can be interpreted as the limits of uncertainty region, but still present a valid output for some other gene regulation network used as an information processing system. Dimer concentrations must have valid interpretations t_{REF} time after the switch or refresh of the state is performed. Both values can be evaluated with the measurements of concentration levels t_{REF} time after the stable state is achieved due to switch or due to refresh of the system's state. Based on the results of the simulations $C_{OH}(\mathrm{min})$ was estimated to 200 nM and $C_{OL}(\mathrm{max})$ to 50 nM. Region of uncertainty can be estimated

to $(50nM, 200nM)$.

5.7 Evaluating the switching times

Switching times can be evaluated in the same way as in semiconductor circuits [17, 18]. We are facing two types of signal switching in digital systems, i.e. from low to high level and from high to low level. Two different switching times need to be evaluated, i.e. *rise time* (t_r) and *fall time* (t_f). Evaluating the switching times on an example of signal observed in our system is presented in Figure 6.

(a)

(b)
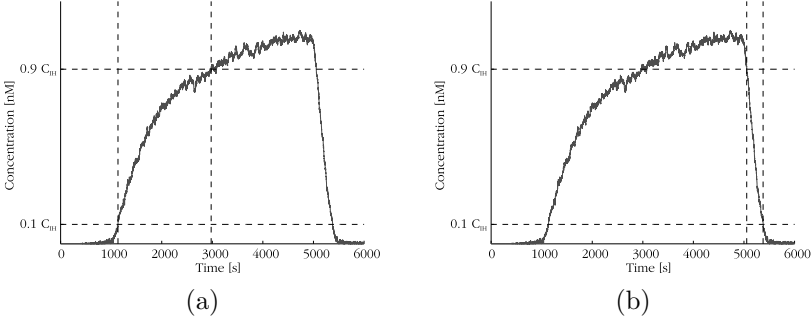
Figure 6. Evaluating the switching times on an example of signal presenting time evolution of protein dimer concentrations. Figure (a) presents the evaluation of rise time and Figure (b) the evaluation of fall time.

5.7.1 Evaluating the rise time

Initially system is set to its stable state, where the observed protein dimer concentration is in the region that can be interpreted as the logical value 0 (low level). Rise time (t_r) can be evaluated as a time needed to perform a switch to its logical value 1 (high level) with the introduction of external influence (i.e. i, respectively j). Time can be accurately measured as time between an event observed protein concentration reaches 10% of $C_{IH}(\text{min})$ and an event it reaches 90% of $C_{IH}(\text{min})$ [17, 18]. Based on the results of the simulations t_r was estimated to 2000 seconds.

5.7.2 Evaluating the fall time

We can evaluate the fall time (t_f) in the similar way as the rise time. We are measuring the time protein dimer concentration needs to fall from 90% of $C_{IH}(\text{min})$ to 10% of $C_{IH}(\text{min})$ [17, 18]. Based on the simulation results t_f was estimated to 400 seconds.

5.7.3 Evaluating the switching time

In order to reach the stable state of the system both of the signals, i.e. both protein dimer concentrations, need to be stabilized after the switch. External

input (i.e. i, respectively j) must be present until both concentrations are within the regions with valid interpretation. Switching time (t_s) can be defined as a maximal value of rise and fall time, i.e. $t_s = \max(t_f, t_r) = 2000$ seconds. External input must therefore be present at least 2000 seconds.

6   Analysis of the modelled system

System's dynamics was analysed from the information processing perspective, i.e. its suitability as a memory element (RS flip-flop biological equivalent) was evaluated. Analysis was performed on the basis of metrics introduced and evaluated in preceding section. Following system's dynamics were observed: *system initialization*, *memorizing the state of the system* and *switching the state of the system*.

6.1   System initialization

Initialization of the system, i.e. establishment of the stable state, was examined. We analysed if the stable state of the system would be established without external inputs, from some undefined initial state of the system, i.e. concentrations of chemical species are arbitrary. Figures 7(a), 7(b) and 7(c) present the time evolution of protein dimer concentrations from three different initial concentrations, that can not be interpreted as a stable state of the system according to Section 5.4. It turns out that external input (i.e. i or j) needs to be introduced temporarily in order to initialize the state of the system similar as in semiconductor memory circuits. Without the initialization system gradually transits to the state that does not have a valid logical interpretation.

Stable state is achieved with the temporal presence of external input. Transition from a state that does not have a valid logical interpretation to a valid stable state is presented in Figure 7(d). External input, i.e. signal i or j, need to be present at least time t_s, which was estimated in Section 5.7. After this time the system is capable of memorizing the state for at least t_{REF} time, which was defined in Section 5.3.

Initialization of the system is performed with the activation of *Set*, respectively *Reset* input. Regarding the scheme of the biological circuit presented in Figure 2 functionalities of these inputs are in accordance with the functionalities of input i, respectively j.

6.2   Memorizing the state of the system

Our hypothesis is that analysed biological system is capable of memorizing its state and can therefore be used as a memory biological circuit. After its initialization or after the switch is performed, state is memorized for at least t_{REF} time, which was determined in Section 5.3. If no other switch occurs within this time, state of the system has to be refreshed with the temporal presence of external signal i or j. Regarding the scheme of the biological circuit presented in Figure 2, state which is defined with high concentration of protein dimer X_2 is refreshed with temporal presence of external signal j and state which is defined with high concentration of protein dimer Y_2 with
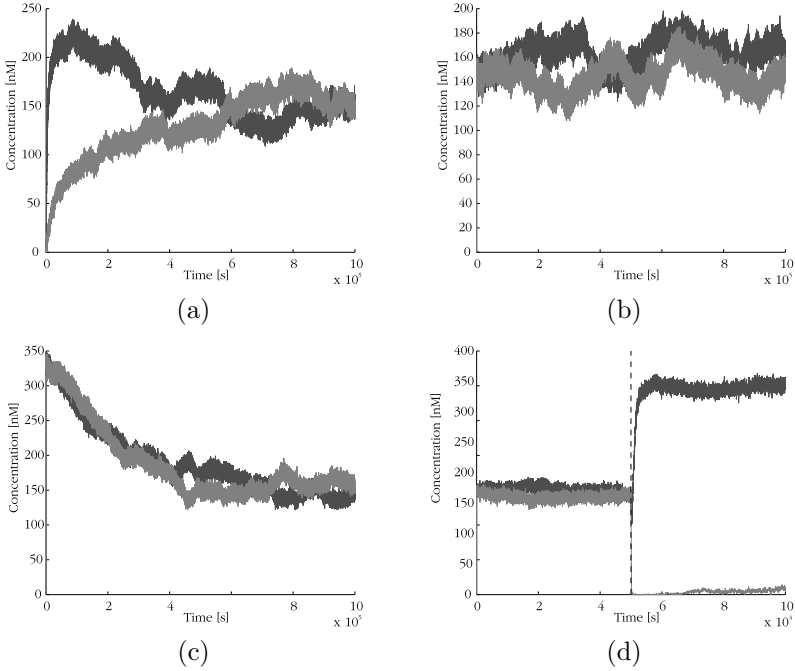
Figure 7. Time evolution of protein dimer concentrations from different initial states of the system. Figures (a), (b) and (c) present a scenario where state that does not have a valid logical interpretation is achieved from some arbitrary initial state. Figure (d) presents initialization of the system state with the introduction of external input, i.e. i or j. Concentrations of protein dimer X_2 are presented with dark-grey line and concentrations of protein dimer Y_2 with light-grey line.

temporal presence of external signal i. In each case, external signal must be present at least t_s time (see Section 5.7). Memorizing the state of the system with refreshing is presented in Figure 4(b).

6.3   Switching the state of the system

Switching of the state is performed similar as refreshing, i.e. with temporal presence of external input i, respectively j, which increases the degradation rate of protein dimer with high concentration. Its concentration drastically decreases and repression of other protein is thus stopped. In other words logical switch is performed. Roles of external inputs i and j are similar as the roles of inputs R and S in RS flip-flop. We presume that external influence presents a binary signal such as light [3]. We can thus describe its presence with only two values, i.e. signal can be absent (0) or present (1). Switching the state of the system is presented in Figure 8, where Figure (a) presents the time evolution of protein dimer X_2 concentration, Figure (b) the time evolution of protein dimer Y_2 concentration, Figure (c) presence of input i and Figure (d) presence of input j.
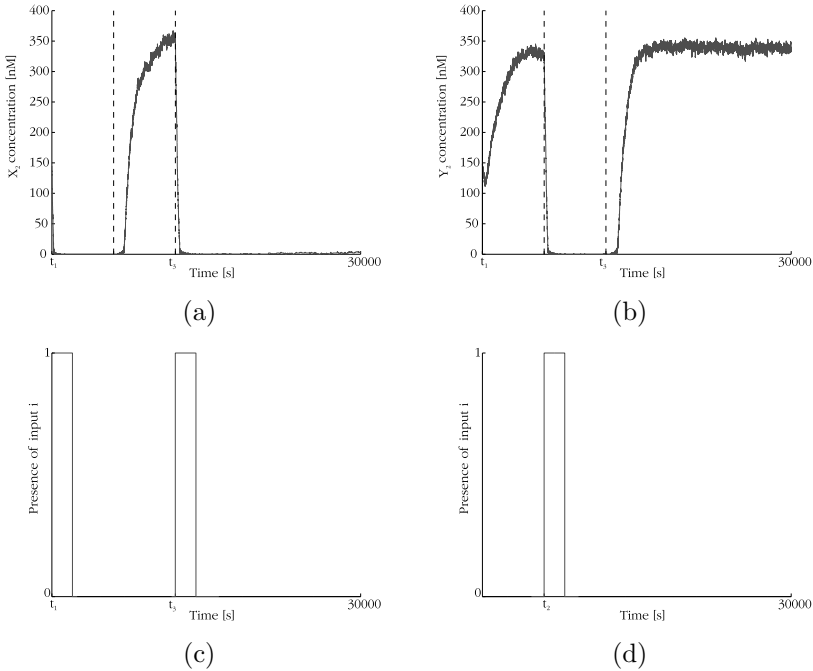
Figure 8. Switching the state of the system with external inputs introduction. Figures (a) and (b) present the courses of protein dimers concentrations (X_2 and Y_2) in dependence of presence of external inputs (i.e. i and j) which is presented in Figures (c) and (d).

Switching times of modelled biological system are obviously much higher than the switching times of semiconductor circuits. Responses of such systems in general can be accelerated for certain amount of time with the selection of different reactions and different chemical species. Noticeable acceleration can also be achieved with employment of massive parallelism which is inherent to observed biological processes [14]. In spite of all described biological systems are still very slow regarding the speed of modern electronic circuits. Since we are not trying to find a direct substitution for current information processing technology but are only trying to transfer its methods to the field of synthetic biology, this should not be treated as a disadvantage of biological systems. Moreover target applications of such systems differ from the ones of modern computers and can be found in the fields of agriculture, energy, pharmacy and biomedicine where response timescales can reach from minutes to hours or even days. In this context processing speed is not an issue.

Conclusions

Introduction of metrics for evaluation of biological systems that are in our opinion appropriate for information processing was presented in the article. The metrics were evaluated on an example model of simple biological system, i.e.

gene regulatory network presenting the toggle switch circuit. Analysis of its dynamics from the information processing perspective on the basis of evaluated metrics was also demonstrated. While the obtained values are strictly dependent on the modelled system, more precisely, on the chemical species and reactions used, the metrics evaluation remains the same for an arbitrary biological system. Moreover the evaluation is practically the same whether it is performed on simulation results of established mathematical models or on experimental data gathered from laboratory experiments. Our main contribution is therefore in the establishment of the metrics and the demonstration of principles of their evaluation. Based on the values of established metrics, an analysis of specific system's dynamics can be performed in order to objectively estimate its suitability as an information processing structure and as a potential building block of a more complex biological processing system.

Acknowledgments

References

[1] Leonard M. Adleman. Computing with DNA. *Scientific American,*, **279**:54–62, 1998.

[2] Uri Alon. *An Introduction to Systems Biology.* Chapman & Hall/CRC, 2007.

[3] Hilary S. Booth, Conrad J. Burden, Markus Hegland and Lucia Santoso. *Mathematical Modeling of Biological Systems, Volume I*, chapter A Stochastic Model of Gene Regulation Using the Chemical Master Equatio, pp. 71–81. Birkhauser Boston, 2007.

[4] Hana El Samad, Mustafa Khammash, Linda Petzold and Dan Gillespie. Stochastic modeling of gene regulatory networks. *International Journal of Robust and Nonlinear Control*, **15**:691–711, 2005.

[5] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**:335–338, 2000.

[6] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, **297**:1183–1186, 2002.

[7] Timothy S. Gardner, Charles R. Cantor and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**:339–342, 2000.

[8] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, **81**:2340–2361, 1977.

[9] Mads Kaern, William J. Blake and J.J. Collins. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, **5**:179–206, 2003.

[10] Mads Kaern, Timothy C. Elston, William J. Blake and James J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews*, **6**:451–464, 2005.

[11] Zvi Kohavi. *Switching and Finite Automata Theory*. McGraw-Hill, Inc., 1978.

[12] Adiel Loinger and Ofer Biham. Stochastic simulations of the repressilator circuit. *Physical Review*, **76**:051917, 2007.

[13] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, **456**:516–520, 2008.

[14] Zoltan Szallasi. Genetic network analysis in light of massively parallel biological data acquisition. In *Proc. Pac. Symp. Biocomput. (PSB99*, pp. 5–16. World Scientific Publishing, 1999.

[15] Tianhai Tian and Kevin Burrage. Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the National Academy of Sciences*, **103**:8372–8377, 2006.

[16] Marcel Tigges, Tatiana T. Marquez-Lago, Jrg Stelling and Martin Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, **457**:309–312, 2009.

[17] Richard F. Tinder. *Engineering Digital Design, Revised Second Edition*. Academic Press, 2000.

[18] John F. Wakerly. *Digital Design: Principles and Practices Package, 4th Edition*. Prentice Hall International, Inc., 2005.

[19] Ron Weiss, George Homsy and Tom Knight. Toward in vivo digital circuits. In *Proceedings of the Dimacs Workshop on Evolution as Computation*, 1999.