Evolved Artificial Signalling Networks for the Control of a Conservative Complex Dynamical System

Luis A. Fuente[1,2], Michael A. Lones[1,2], Alexander P. Turner[1,2], Susan Stepney[1,3], Leo S. Caves[1,4], Andy M. Tyrrell[1,2]

[1] Department of Electronics {laf509, mal503, apt503, amt}@ohm.york.ac.uk
[2] York Centre for Complex Systems Analysis (YCCSA)
[3] Department of Computer Science susan.stepney@cs.york.ac.uk
[4] Department of Biology lsdc1@york.ac.uk
University of York, Heslington, York, YO10 5DD, UK

Abstract. Artificial Signalling Networks (ASNs) are computational models inspired by cellular signalling processes that interpret environmental information. This paper introduces an ASN-based approach to controlling chaotic dynamics in discrete dynamical systems, which are representative of complex behaviours which occur in the real world. Considering the main biological interpretations of signalling pathways, two ASN models are developed. They highlight how pathways' complex behavioural dynamics can be captured and represented within evolutionary algorithms. In addition, the regulatory capacity of the major regulatory functions within living organisms is also explored. The results highlight the importance of the representation to model signalling pathway behaviours and reveal that the inclusion of crosstalk positively affects the performance of the model.

1 Introduction

Cells need to engage in many forms of communication in order to sense and respond to the outside world. This capacity is vital for cells to survive. Cellular signalling involves a chain of events that permits cells to interact with their environment. It starts with the triggering of a biochemical signal and terminates with an adaptive cellular response. Classically, cellular signalling may be conceived as follows: a surface receptor binds an extracellular messenger (e.g. hormone, growth factor), which diffuses an intracellular signal to an effector protein inside the cell. This then produces secondary messengers, which transmit the information further into the cell. Spatially or temporally variable catalytic reactions or cascades of protein kinases finally lead to changes in gene expression, bringing about a change in cellular activity.

In this paper we propose a new Artificial Signalling Network (ASN) approach to modelling the spatial properties and temporal topologies of cellular signalling, capturing its intrinsic dynamics. In order to test the model we apply it to the

control of a numerical dynamical system, whose properties mirror the complexity of cellular environments. Controlling dynamics also represents a classical multi-disciplinary problem in its own right.

This paper is organised as follows: Section 2 presents a brief overview of dynamical systems, Section 3 reviews the current literature on ASNs, Section 4 introduces the new model and defines its methodology, Section 5 presents some initial results, and Section 6 concludes.

2 Dynamical Systems

A dynamical system is a mathematical model where a function, or *evolution rule*, characterises its state based on the system's current state and initial conditions [12]. The evolution rule defines the motion and behaviour of the system across the state space. Dynamical systems are initially divided into *autonomous* and *non-autonomous*. The former is a closed system whose dynamics are not perturbed by the outside world. The latter defines an open system changing over time, as inputs are received from an external environment. Likewise, dynamical systems can be *discrete* or *continuous* in time, depending on the type of evolution rule: difference equations in the former and differential equations in the latter.

Given a set of initial points within a discrete state space, the evolution rule defines their *trajectories* as a sequence of states over a period of time. A dynamical system where trajectories do not contract to a limited region of the state space is known as a *conservative* system.

Dynamical systems can display a wide range of behaviours. The most interesting are those involving holistic irregular and unpredictable properties; this atypical dynamism is known as *chaos*. Despite being deterministic, chaotic systems display aperiodic behaviours characterised by an exponential sensitivity to initial conditions and the existence of strange attractors. Whereas the former suggests that small changes in the initial conditions convey highly different trajectories throughout the state space, the latter defines fractal and non-linear regions where trajectories may converge.

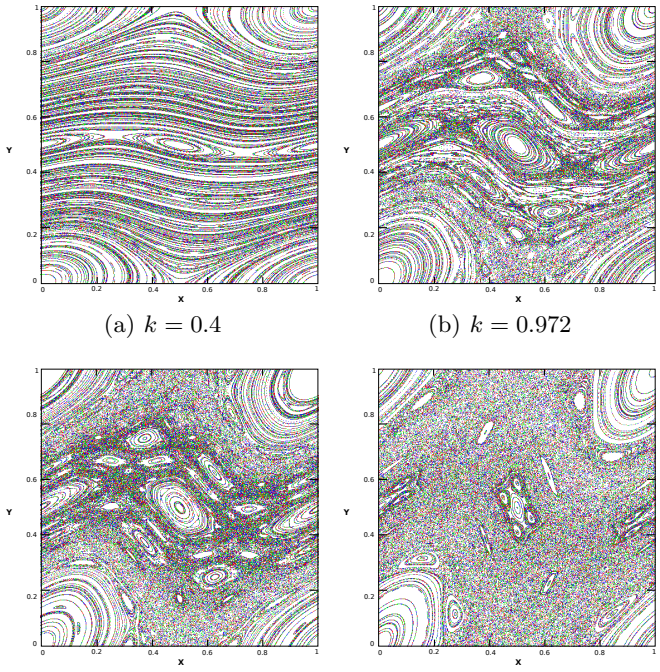
2.1 Chirikov's Standard Map

Chirikov's standard map [4] is a conservative and discrete two-dimensional dynamical system representing iteratively the interactions of two canonical variables within the unit square:

$$x_{n+1} = (x_n + y_{n+1}) \bmod 1 \qquad y_{n+1} = y_n - \frac{k}{2\pi}\sin(2\pi x_n) \tag{1}$$

One of the map's main properties is its capacity to represent different dynamics as its nonlinearily increases. Thus, low values of k preserve an ordered state where trajectories lead to periodic and quasi-periodic trajectories bounded on the y-axis (see Fig.1(a)). As k increases, chaotic dynamics arise in the form of chaotic islands along the y-axis, which are never visited (see Fig.1(b)). The type

of trajectories depends on the map's initial conditions. The map shows a behavioural inflection point, k_c, at $k \approx 0.972$. Initial impermeability progressively disappears as $k > k_c$ (see Fig. 1(c)–(d)), enabling trajectories to vertically travel across the map. The example in Figure 1 shows the permeability of the map increasing as k increases, characterised by the gradual encroachment of the chaotic regions.



(a) $k = 0.4$ (b) $k = 0.972$

(c) $k = 1.1$ (d) $k = 2.0$

Fig. 1. Sampled trajectories of Chirikov's standard map using different values of k, showing the change from the ordered to the chaotic state. Each map is plotted using 400 randomly chosen initial points across the unit interval over 800 iterations.

2.2 State Space Targeting

The sensitivity underlying chaotic dynamics brings about a broad band of complex, unstable and unpredictable behaviours using arbitrary control signals. *Controlling chaos* or *chaos targeting* attempts to provoke large behavioural changes in the dynamics using relatively small perturbations, which are achievable by the modification of the system's control signals. Therefore, it looks at finding the fastest path from an initial condition to a target point. Existing research has shown that targeting in conservative systems, such as Chirikov's standard

map, is achievable using small perturbations to drive across the different chaotic regions of the state space [9, 14]. Consequently, the map becomes navigable from the bottom to the top and it is possible to find a controller able to transverse it.

3 Artificial Signalling Networks

As an abstraction of cellular signalling, Artificial Signalling Networks try to model the particular characteristics that allow cells to take chemical signals as inputs and generate some adaptive output. Given our motivation to preserve biological plausibility, we are interested in investigating the ASN's ability to implement control functionalities. From a computational perspective, the importance of ASNs lies in their capacity to coordinate the set of events within cells that trigger robust, efficient and specific responses, their ability to work as independent processing units and their capacity to adapt to environmental perturbations.

One way to model ASNs relies on the quantitative description of particular pathways. Experimental and mathematical approaches facilitate the identification of the functional elements as well as their interactions in concrete pathways, thereby simplifying their modelling. Thus, the fuzzy model suggested in [7] computes the dynamics of the IL-6 pathway based on the state of the components, the initial inputs and a set of fuzzy rules. Likewise, the validity of logic-based modelling has been widely demonstrated [10]. These models have a direct physical basis. Said et al. in [13] take a more abstract approach, modelling the interaction between two participating elements, and then, applying it to simulate MAP kinase cascade as a Markov chain. However, the reconstruction of these pathways was insufficiently accurate since the complexity displayed by some of the components could not be captured.

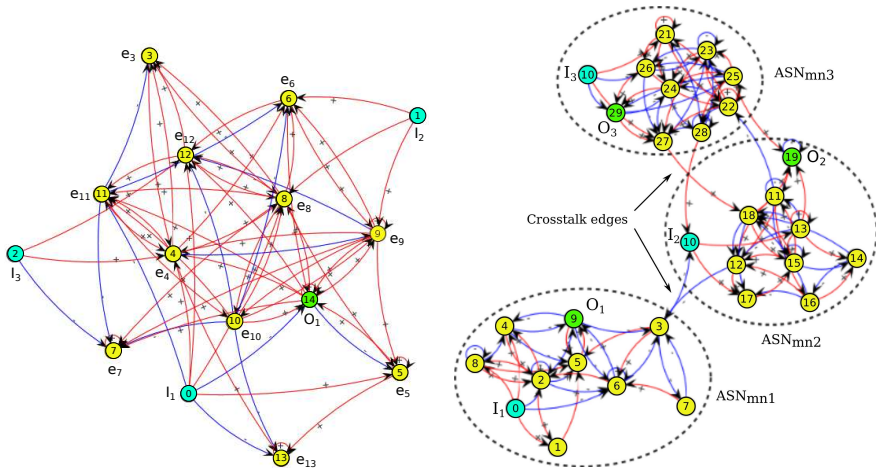
Another way to design ASNs is to use evolutionary algorithms. They can induce complex behaviours in a concise and evolvable way [9] and some specific functionalities are achievable only through evolutionary processes [6]. In fact, evolved ASNs have been successfully used to capture simple forms of biological signal processing [3, 5]. In this paper we propose an alternative approach: we use a generic evolved artificial signalling network, where no specific information of either the participating elements or their interactions is needed. Therefore, all limitations emerging from the pathway's particular characteristics are dismissed and the ASN's topology is the result of its interaction with the environment. This increases its adaptability when facing different types of environment. A similar approach has been suggested in [8]. However, it models ASNs as Boolean networks and limits the connectivity between the participating elements.

4 State Space Targeting with ASNs

Most of the signalling processes inside the cells involve complex interactions between enzymes. Although these interactions may vary in size, they are essential in the transmission of signals. In practice, enzymes are not functional unless they

are grouped together into a biological structure. Likewise, some of the main cellular functions are only achievable under certain spatial distributions. There are a wide variety of abstractions aiming to represent the properties of intracellular signalling networks. However, many of them fail to fulfill this objective, or it is only partially achieved. For example, Bayesian Networks [11, 15], limit the representation of complex dynamics due to their acyclic nature. This paper proposes the usage of *interaction graphs* to capture the topological and temporal patterns intrinsic to signalling pathways. An interaction graph is a mathematical representation modelling the dynamical behaviours of a system formed by multiple actors interacting over time; thus, we consider ASNs as dynamical systems defined by interactions between enzymes.

According to the different types of pathways inside cells, we introduce two approaches for ASN modelling. Whilst the first model considers ASNs as a subtype of metabolic network (see Fig. 2(a)), the second considers them to be cascades of protein kinases (see Fig. 2(b)). Both approaches extend the model described in [9]. Both are continuous-valued models as this enables a more realistic representation of biological systems. To allow a valid comparison between both models, they are deterministic and synchronous.



(a) $\mathrm{ASN_{mn}}$ (b) $\mathrm{ASN_{c}}$

Fig. 2. Representation of both ASN models. Each models has three inputs, $\mathrm{I_i}$, and one global output $\mathrm{O_i}$. $\mathrm{ASN_{mn}}$ contains 15 enzymes. $\mathrm{ASN_{c}}$ has 3 $\mathrm{ASN_{mn}}$, each of which has 10 enzymes. Crosstalk edges are the arcs connecting two $\mathrm{ASN_{mn}}$s.

The artificial signalling network as a metabolic network ($\mathrm{ASN_{mn}}$) is defined as a directed interaction graph, where the nodes are an indexed set of enzymes and the edges represent their biochemical reactions. Every en-

zyme contains a set of substrates, a set of products and a mapping function relating the concentrations of both sets of chemicals. Formally: $\text{ASN}_{\text{mn}} =< C, E, R, I_E, O_E >$, where:

- $\boldsymbol{C}$ is the indexed set of random chemical concentrations $\{c_0, c_1, \ldots, c_n : \mathbb{R}\}$.
- $\boldsymbol{E}$ is the indexed set of enzymes $\{e_0, e_1, \ldots, e_n \ : \ e_i =< S_i, P_i, m_i >\}$, where:
  - $\boldsymbol{S_i} \subseteq \boldsymbol{C}$ is the concentration of the substrates used by the enzyme.
  - $\boldsymbol{P_i} \subseteq \boldsymbol{C}$ is the concentration of the products generated by the enzyme.
  - $\boldsymbol{m_i} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the enzymes' substrate-product mapping function.
- $\boldsymbol{R}$ defines the set of enzymatic reactions $\{r_0, r_1, \ldots, r_n \ : \ r_i \in \{+, -\}\}$. Negative and positive values indicate enhancing and inhibition respectively.
- $\boldsymbol{I_E} \subset \boldsymbol{E}$ is the set of enzymes used as inputs.
- $\boldsymbol{O_E} \subset \boldsymbol{E}$ is the set of enzymes used as outputs.

The execution of the ASN proceeds as follows:

1. S_i and P_i are randomly initialised (if ASN not previously executed).
2. The concentrations of S_i in I_E are set by the external inputs.
3. At each time step, each enzyme e_i applies its mapping function m_i to determine the new concentration of the products P_i based on the concentration of its substrates S_i. In the particular case where the substrate is provided by multiple enzymes, the new concentration is the mean output of all different contributing enzymes.
4. After a number of time steps the execution is halted and the concentrations of the products in O_E are copied to the external outputs.

The artificial signalling network as a set of protein kinases cascade (ASN_{C}) extends the previous model by grouping the enzymes into an indexed set of ASN_{mn}, each of which represents a signalling pathway. Additionally, pathway crosstalk is simulated by a set of edges connecting two ASN_{mn}s. Formally $\text{ASN}_{\text{C}} =< ASN, C_R, O_E, I_E >$, where:

- $\boldsymbol{ASN}$ is an indexed set of artificial signalling networks $\{asn_0, asn_1, \ldots, asn_n \ : \ asn_i \equiv ASN_{mn} =< C, E, R, I_E, O_E >\}$.
- $\boldsymbol{I_E} \subset \boldsymbol{E}$ is the set of enzymes used as inputs, where $|ASN_{mn}| = |I_E|$.
- $\boldsymbol{O_E} \subset \boldsymbol{E}$ is the set of enzymes used as outputs.
- $\boldsymbol{C_R} \in [0, 1]$ is the probability of crosstalk.

The execution of ASN_{C} is similar to that of ASN_{mn}:

1. S_i and P_i are randomly initialised (if ASN not previously executed).
2. The concentrations of S_i in I_E are set by the external inputs. Each ASN_{mn} has only one input.
3. At each time step, each enzyme e_i applies its mapping function m_i to determine the new concentration of the products P_i based on the concentration of its substrates S_i. When the substrate is provided by multiple enzymes, the new concentration is the mean output of all different contributing enzymes. Additionally, enzymes having a crosstalk edge have their products asymptotically reduced to half of their maximum rate.

4. After a number of time steps ($t_s \in [1, 100]$) the execution is halted and the external outputs are calculated as the mean output value of all contributing networks.

We also analyse the effect of having two types of enzymes depending on the number of times they are phosphorylated (single and double). This more closely represents the different phosphorylation states in cascades of protein kinases.

4.1 Mappings

Three types of parameterisable functions are chosen as enzyme mappings: the Hill, the Michaelis-Menten and the first-order kinetics equations. They are the most common models of molecular regulatory functions within living organisms.

The Hill equation describes the cooperative level between an enzyme and its substrate as $f(x) = v|x|^h/k^h|x|^h$, where $v \in [0, 1]$ is the asymptotic threshold, $k \in [0, 1]$ determines its gradient and $h \in \mathbb{R}^n$ is the hill coefficient indicating the degree of cooperativeness. The equation can also be extended by adding the probability of binding $\beta \in [0, 1]$. If $h = 1$ the Hill equation is equivalent to the Michaelis-Menten equation. For multiple inputs, $x = \sum_{j=0}^{n} \frac{i_j w_j}{n}$. Negative values indicate inhibition and $f(x)^- = 1 - f(x)$.

The Michaelis-Menten equation characterises the enzyme kinetic reactions. It is a hyperbolic function $f(x) = v|x|/(k - |x|)$, where $v \in [0, 1]$ is the asymptotic threshold and $k \in [0, 1]$ determines its gradient. For multiple inputs, $x = \sum_{j=0}^{n} \frac{i_j w_j}{n}$, where $i_0 \dots i_n$ are inputs and $w_0 \dots w_n \in [-1, 1]$ are the corresponding input weights. Negative values indicate inhibition and $f(x)^- = 1 - f(x)$.

The Multi-Dimensional Michaelis-Menten equation defines the enzymes' kinetics when substrates are produced by multiple enzymes based on the probability of binding as $f(x) = \sum_{i=0}^{k} \beta_i (x_i/k_i)^{n_i} / 1 + \sum_{i=0}^{n} (x_i/k_i)^{m_i}$, where $\beta \in [0, 1]$ is the binding probability, $v \in [0, 1]$ is the asymptotic threshold and $k \in [0, 1]$ determines its gradient. $m, n \in \mathbb{R}^n$, where $m = n$ for activation and $n = 0$ and $m > 0$ for repression [1].

The first-order kinetics equation is the simplest kinetics model relating to the rate of phosphorylation of an enzyme to the concentration of its active site and the concentration of the unphosphorylated substrate. When single phosphorylated $f(x) = v|x|/(1 + |x|)$, and when double $f(x) = vx^2/(1 + |x| + x^2)$, where $v \in [0, 1]$ is the asymptotic threshold. For multiple inputs, $x = \sum_{j=0}^{n} \frac{i_j w_j}{n}$. Negative values indicate inhibition and $f(x)^- = 1 - f(x)$.

4.2 Methodology

Both ASN models have been evolved using a standard generational evolutionary algorithm with tournament selection (size=4), uniform crossover (rate = 0.48), and point mutation (rate = 0.16). 40% of the solutions are mutated. An ASN is encoded as an array of genes, followed by an integer within the interval [0, 100] representing the number of time steps for execution. Crossover points lie between the enzymes' boundaries. In an attempt to reduce the complexity of the analysis, the number of genes and enzymes has been fixed at 10. All runs terminate after 100 generations.

Initial chemical values and mapping parameters are represented using floating-point values and mutated using a Gaussian function with its center at the current values. However, mutation is constrained to one of the following operations to fulfill the restrictions proposed in [16] to model the reactions between molecules:

1. Increasing or decreasing the chemical values.
2. Changing the state of the biochemical reactions by modifying the parameters of the mapping functions.
3. Variation of the reaction rates by changing their weight values.
4. Adding or removing participants (edges) to/from the reactions.
5. Variation of the number of time steps.

Each ASN is represented as an interaction graph in which each vertex is an enzyme. External inputs, which represent the controller's state space location at the start of execution, are always delivered to the inputs of low-numbered enzymes (in terms of the network's genetic encoding). External outputs, which determine the new values for control signals, are always read from the outputs of high-numbered enzymes.

Traversing Chirikov's standard map: The goal is to evolve an ASN-based controller which can guide trajectories from a designated region at the bottom of the map to a designated region at the top of the map by modulating the control signal k within the range $[1.0, 1.1]$. Inputs of the ASN are the position in the map and the Euclidean distance from the current position to the top-centre of the map, $< x, y, d >$, the output is the value of k (suitably scaled). The fitness function is the number of steps the controller needs to transverse the map and is limited to a maximum number of 1000 steps. Controllers exceeding this threshold are penalised with a fitness of 2000 steps. A population size of 500 is used.

5 Results

Results from controlling Chirikov's standard map using both ASN approaches are shown in Fig. 3. Both models led to effective controllers which were able to solve the problem (see Fig. 5(a)–(b)). The best performance comes from $\mathrm{ASN}_{\mathrm{mn}}$, but $\mathrm{ASN}_{\mathrm{c}}$ can also lead to valid solutions when every pathway computes its dynamics independently or quasi-independently. The degree of crosstalk has a

significant effect upon the solutions (see Fig. 4): low crosstalk seems to be beneficial, whereas high values add uncorrelated noise reducing the overall system behaviour. Similar conclusions on the effect of crosstalk were noted in [2].

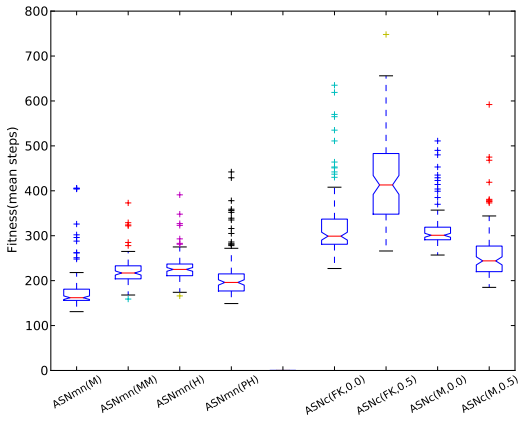


Fig. 3. State space targeting results using evolved ASNs with (M)ichaelis-Menten, (M)ulti-Dimensional (M)ichaelis-Menten, (H)ill, (P)robabilistic (H)ill, and (F)irst-order (K)inetics as regulatory equations. For ASN_C, the value next to the type of equation indicates the crosstalk rate. Summary statistics of the 100 runs are shown as box plots. Low values are better.

Perhaps the most interesting result is the capacity of ASN_C to solve the problem, even where there is no crosstalk. There is certainly an indication from the results that complex problems can be divided into smaller and independent tasks, which individually lead to valid solutions. A similar organization can be seen inside cells, which not only enclose a wide number of specific pathways, but also limit their interactions by using compartments. It highlights the essential role that crosstalk may play in the formation of more complex and realistic models. Despite the validity of the results, we believe that the procedure used to determine the ASN_C output (mean of all contributing networks) is not biologically plausible and therefore we might be losing some useful information. We hypothesise that ASN_C would be better able to cope with incomplete or corrupt environmental information, enabling better environmental adaptation. This is something we aim to test in future work.

The choice of the regulatory function also has significant consequences. Generally, all regulatory functions work better for the ASN_{mn} model. The Michaelis-Menten equation seems to provide the most effective controllers in both approaches, however it needs to be complemented with a relatively high crosstalk probability in ASN_C. This effect contrasts with the results obtained using the

first-order kinetics equation, which offer a lower performance and require a low crosstalk rate to achieve similar results (see Fig. 4).

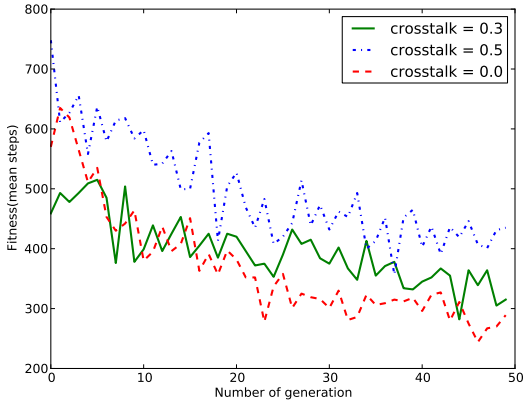


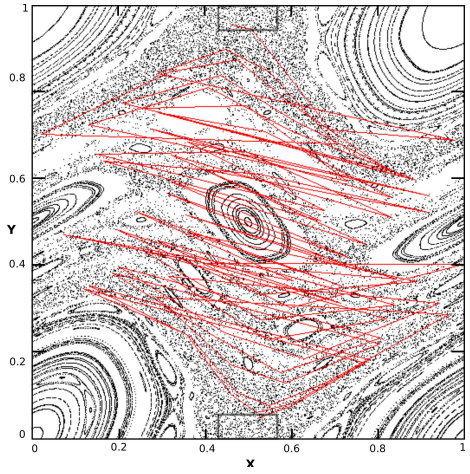
Fig. 4. Effect of crosstalk rate upon the effectiveness of ASN_C controllers, showing the rate of change in fitness 50 generations averaged over 100 runs.

6 Conclusions

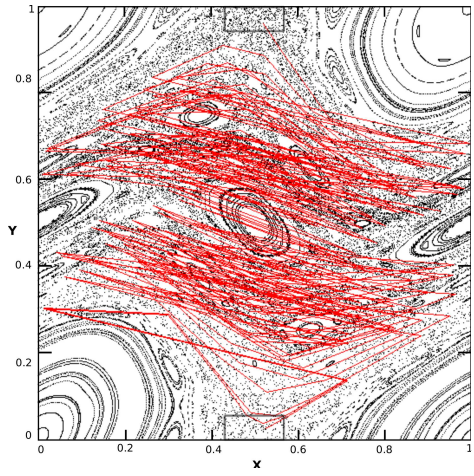
In this paper we have presented an approach to modelling signalling pathways using evolutionary algorithms. Our results are encouraging; demonstrating that evolved artificial signalling networks can be used to regulate complex dynamical behaviours within Chirikov's standard map. Notably, our results show that effective controllers can also be found when signalling networks are interpreted as sets of either pathways or cascades of protein kinases. These results are broadly similar to the AGN- (Artificial Genetic Network) and AMN- (Artificial Metabolic Network) based controllers described in [9]. This verifies that the robustness and adaptability of signalling networks can be evolved. Likewise, an accurate representation of their spatial and temporal dynamical properties was achieved with no additional knowledge of the surrounding environment and the participating elements.

This paper has highlighted the importance that representation plays in effectively modelling signalling pathway dynamics. Our results show how different ASN models may be suited to different problems. In particular, we have illustrated the sensitivity of ASNs to the level of crosstalk between pathways, which has a large impact upon the effectiveness of the controllers. However, we believe crosstalk will be an important mechanism when looking for more complex and sophisticated representations of signalling networks.

In future work, we plan to explore how ASNs' complex dynamical behaviours can be affected by their spatial and temporal properties, and to look at how these can be used to solve more complex real-word problems.



(a) ASN_{mn}



(b) ASN_{c}

Fig. 5. Example of state space targeting from an area at the bottom, $(0.45, 0) \rightarrow (0.55, 0.05)$, to a region at the top $(0.45, 0.95) \rightarrow (0.55, 1)$, in Chirikov's Standard Map with both ASN models. The ASN_{mn} moves from the bottom to the top in 91 steps. The ASN_{c} does it in 262 steps. The plotted standard map corresponds to the k value at the end of the ASN's execution.

Acknowledgements

This research is supported by the EPSRC ref: (EP/F060041/1), Artificial Biochemical Networks: Computational Models and Architectures.

References

1. Alon U.: An introduction to systems biology: design principles of biological circuits. Chapman & Hall/CRC Mathematical and Computing Biology Series (2007)
2. Arias, M.A., Hayward, P.: Filtering transcriptional noise during development: concept and mechanisms. Nature Reviews Genetics **7(1)** (2006) 34–44
3. Bray, D., Lay, S.: Computer simulated evolution of a network of cell signalling molecules. Biophysical Journal, **66** (1994) 972–977
4. Chirikov, B.V.: Research concerning the theory of nonlinear resonance and stochasticity. Tech. rep., Institute of Nuclear Physics, Novosibirsk (1962)
5. Deckard, A., Sauro, M.H.: Preliminary studies on the *in silico* evolution of biochemical networks. Chembiochem, **5(10)** (2004) 1423–1431
6. Decraene, J., Mitchell, G.G., McMullin, B.: Evolving artificial cell signalling networks: Perspectives and Methods. Studies in Computational Intelligence (SCI) **6** (2009) 167–186
7. Huang, Z., Hahm,J.: Fuzzy modeling of signal transduction networks Proc, 17th World Control, The International Federation of Automatic Control (2008) 15867–15872
8. Klamt, S., Saez-Rodriguez, J., Lindquist, J.A., Simeoni, L., Giles, D.E.: A methodology for the structural and functional analysis of signalling and regulatory networks. BMC Bioinformatics **7(56)** (2006) 1–26
9. Lones, M.A, Tyrrell, A.M, Stepney, S, Caves, S.L.: Controlling complex dynamics with artificial biochemical networks. Proc. 2010 European Conference on Genetic Programming, Lecture Notes in Computer Science **6021**, (2010) 159–170
10. Morris, K.M., Saez-Rodriguez, J., Sorger, K.P, Lauffenburger, A.D: Logic-based models for analysis of cell signalling networks. Biochemistry **4(49)** (2010) 3216–3224
11. Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., Lauffenburger, D.A.: Bayesian networks approach to the cell signalling pathway modeling. Science's STKE **148** (2002) 38–42
12. Stepney, S.: Nonclassical computation: a dynamical systems perspective. In: Rozenberg, G., Bäck, T., Kok, N.J. (eds.) Handbook of natural computing vol. 2 chap. 52 (2011) Springer
13. Said, M. R., Oppenheim, A. V., Lauffenburger, D. A.: Modelling cellular signal processing using interacting Markov chains, Proc. International Conference on Acoustic, Speech, Signal Processing (ICASSP 2003), Hong Kong 41–44
14. Schroer, C.G., Ott, E.: Targeting in Hamiltonian systems that have mixed regular/chaotic phase spaces. Chaos, **7** (1997) 512–519
15. Tulupyev, A., Nikolenko, S.: Directed cycles in Bayesian belief networks: probabilistic semantics and consistency checking complexity. MICAI 2005: Advances in Artificial Intelligence, (2005), 214–223
16. Ziegler, J., Banzhaf, W.: Evolving control metabolisms for a robot. Artificial Life, **7** (2001) 171–190