

# **Towards Algorithms for Autonomous Experimentation**

C. J. Lovell and K.-P. Zauner

School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK  
{cjl07r,kpz}@ecs.soton.ac.uk

## **1 Introduction**

Modelling biological systems, is impaired by the cost of experimentally obtaining the data required to build the models. The resources available are typically very limited compared to large experimental parameter spaces, so have to be used efficiently. Similarly, in engineering with biological systems such as is found in synthetic biology and molecular computation, models of the phenomena to be harnessed are required. These models can only be obtained experimentally. As a consequence, for engineering with biological systems to become more prevalent, tools and techniques are required to assist in the experimentation performed to develop these models. Here we focus on one such tool, a computational system capable of autonomously investigating an experimental parameter space to identify phenomena that exist within. We call this computational system an autonomous experimentation system.

Autonomous experimentation systems try to capture the efficiency of experimentalists, who are able to successfully navigate a seemingly boundless space of potential experiments. Autonomous experimentation systems develop hypotheses, plan experiments and perform experiments, in a closed loop manner without human interaction. As a foundation for our approach to autonomous experimentation, we draw on ideas from the philosophy of science.

## **2 Experimentation**

Experimentation should work to disprove hypotheses [1]. A hypothesis gives a possible explanation for some observed phenomena. Hypothesis lead experimentation can benefit from considering many hypotheses simultaneously, so as to allow for different explanations for a particular phenomenon to be developed without prejudice [2, 1]. While human scientists are limited in the number of different working hypotheses that they can realistically contemplate and visualise at any one time, a computer has no such limitation and could compare many thousand different hypotheses simultaneously.

The hypotheses that are developed through experimentation, need not be mechanistic in nature. The investigation of the relationship between cause and effect can be performed experimentally, without developing mechanistic hypotheses. It is known for instance that children display an ability to associate cause with effect from an early age through their early play [3]. The use of cause and effect experimentation in scientific, medical and engineering work, has allowed for developments in these areas without an understanding for why the cause and effect are related. For example, a cure for scurvy was produced long before anyone understood why the cure worked [4].

### 3 From Experimentation to Autonomous Experimentation

Computational systems capable of scientific discovery are interlinked with artificial intelligence systems. Computational scientific discovery puts into practice artificial intelligence methods and brings real world benefits with it. It is important to note that neither artificial intelligence or autonomous experimentation systems will be able to match the abilities of human creativity. The KEKADA system [5], was an early computational experimentation system that was able to develop hypotheses and plan experiments to investigate an experimental parameter space. The KEKADA system took the approach of performing a broad search of the experimental parameter space through experimentation until a surprising phenomena was found, at which point the system performs more focussed experiments on the surprising phenomena. The hypotheses developed were mechanistic in nature and the system was able to rediscover known phenomena, such as determining the mechanism for how urea is synthesised in the body, and determining the structure of common alcohol, which also showed the systems generality in its hypotheses [5]. The KEKADA systems limitations came to the fore when comparing it to scientists, where scientists have the advantage of being able to employ additional heuristics and so able to solve a significantly larger number of problems. The authors concluded that improvements to the KEKADA system could come through increasing the amount of domain knowledge available to the system.

Domain knowledge to developing hypotheses in other computational experimentation systems. It was shown that reaction pathways could be obtained automatically from experimental evidence and domain knowledge [6]. The use of increased domain knowledge was influential in the DENDRAL project, which built an expert system for use in scientific reasoning, and in particular aiding structure elucidation in organic chemistry [7]. The combination of domain knowledge and experimental results through abductive reasoning, was shown to be able to rediscover the function of genes in known roles [8]. The Robot Scientist project largely automated the majority of the physical side of experimentation and was also able to show that the computational system and human scientists could comparatively well interpret experimental results, albeit in the limited domain space in which the Robot Scientist operated [8].

In the technical application of enzymes for molecular computing, such large amounts of domain knowledge do not exist. A different approach for autonomous experimentation therefore has to be taken than those that already exist. One autonomous experimentation system has moved away from the requirements of large amounts of domain knowledge and instead concentrated on searching the experimental parameter space for surprising phenomena, through a fully closed-loop system [9]. Such an approach does not lend itself well to the development of mechanistic hypotheses, which is why we look to investigate causality in our approach. By investigating causality, we are still able to develop hypotheses that are both useful to a scientist and are free from large amounts of domain information.

### 4 Autonomous Experimentation for Molecular Computing

Enzymes can be thought of as a pattern recognizer working at the molecular level. An enzymes ability to process multiple inputs simultaneously, make it a candidate for use in molecular computing [10]. At present enzymes cannot be designed for purpose, therefore experimentation is necessary to develop these models. An example of a typical experiment that our system could perform, is monitoring the activity of an enzyme by spectroscopically measuring the ultra-violet (UV) absorbance of the enzyme. Models can then be made of the interplay between different substances that effect the enzymes activity and the quantities of those substances, with the resulting UV absorbance. Of particular interest are regions of the parameter space where a combination of factors cause a specific change in the enzymes behaviour. Such changes in behaviour are the properties that could be harnessed in molecular computation. Our system will determine these regions of interest for molecular computation.

A microfluidic device is under development to allow for a fully autonomous, closed-loop experimentation system. This lab-on-a-chip technology uses only small amounts of chemistry for each experiment, which will make autonomous experimentation more practical in terms of cost. We follow a conceptual design similar to those of previous work [5, 11], as shown in figure 1, where hypotheses are generated, experiments are proposed so as to try and disprove hypotheses, then an automatically chosen experiment is performed, with the results being used to update the working set of hypotheses.

Any practical experimentation system has to robustly model a set of data, capable of handling noise on the dependent and independent parameters. Our prototype implementation shows that smoothing splines offer these properties for modelling the type of data we expect. We are currently investigating the use of smoothing splines modified with a Bayesian framework, to develop models with lower dimensional data sets than those we expect to be using in later work. These lower dimensional data sets will allow for easier conceptualisation of the problems of

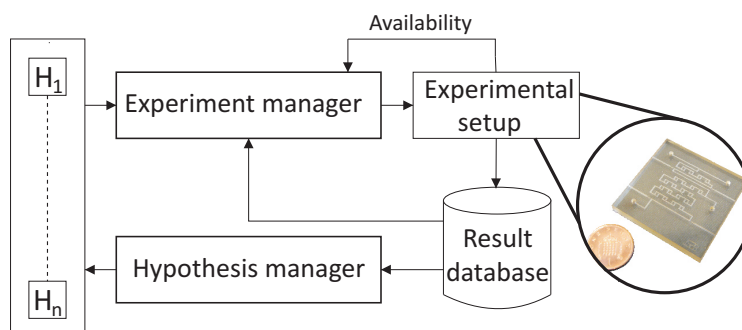


Figure 1: Interplay of experiment manager and hypothesis manager. The set of working hypotheses,  $H_1 \dots H_n$ , are developed and maintained by the hypothesis manager. The experiment manager evaluates the working hypotheses and previous experiment observations, to propose an experiment to perform. The experiment is performed by the experimental setup. A prototype of the microfluidic device that will form the experimental setup is shown.

autonomously developing hypotheses and methods for exploring the experimental parameter space.

## Acknowledgements

The reported work was supported in part by a Microsoft Research European Fellowship to KPZ.

## References

- [1] Platt, J. R. Strong inference, certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* **146** (1964).
- [2] Chamberlin, T. C. The method of multiple working hypotheses. *Science (old series)* **15**, 92–96 (1890). Reprinted in: *Science*, v. 148, p. 754–759, May 1965.
- [3] Gopnik, A., Sobel, D. M., Schulz, L. E. & Glymour, C. Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology* **37**, 620–629 (2001).
- [4] Claridge, J. A. & Fabian, T. C. History and development of evidence-based medicine. *World Journal of Surgery* **29**, 547–553 (2005).
- [5] Kulkarni, D. & Simon, H. A. Experimentation in machine discovery. In Shrager, J. & Langley, P. (eds.) *Computational Models of Scientific Discovery and Theory Formation*, 255–273 (Morgan Kaufmann Publishers, San Mateo, CA, 1990).
- [6] Valdés-Pérez, R. E. Conjecturing hidden entities via simplicity and conservation laws: Machine discovery in chemistry. *Artificial Intelligence* **65**, 247–280 (1994).
- [7] Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A. & Lederberg, J. DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence* **61**, 209–261 (1993).
- [8] King, R. D. *et al.* Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).
- [9] Matsumaru, N., Centler, F., Zauner, K.-P. & Dittrich, P. Self-adaptive scouting—autonomous experimentation for systems biology. In Raidl, G. R. *et al.* (eds.) *Applications of Evolutionary Computing: EvoWorkshops 2004*, vol. 3005 of *Lecture Notes in Artificial Intelligence*, 52–62 (2004).
- [10] Zauner, K.-P. & Conrad, M. Enzymatic computing. *Biotechnol. Prog.* **17**, 553–559 (2001).
- [11] Whelan, K. E. & King, R. D. Intelligent software for laboratory automation. *Trends in Biotechnology* **22** (2004).