Autonomous Experimentation Methods for Characterising Molecular Computing Substrates

Chris Lovell and Klaus-Peter Zauner

Science and Engineering of Natural Systems Group
School of Electronics and Computer Science
University of Southampton, United Kingdom
{cjl07r, kpz}@ecs.soton.ac.uk

Today's computing technology is built on a narrow foundation of materials. Nature demonstrates efficient information processing implemented with macromolecular computing substrates. Molecular materials possess a variety of properties that make them attractive for future computing technologies. However, traditional engineering and design approaches are ill suited to the complexity of these molecules. New methods are required to add macromolecules to the toolkit of the computer engineer. A crucial step towards this goal is the characterisation of the behaviour of macromolecules in the context of other molecules. The present project investigates and develops methods for efficiently characterising molecular materials through computer controlled experimentation. The ultimate aim of this work are algorithms that emulate the experimentation strategies of human experimenters.

Autonomous Experimentation

Experimentation is the process of performing experiments to acquire information that provides more detailed knowledge about some system. Given a set of observations from experiments, a scientist will formulate a number of working hypotheses that try to explain the observations. An experiment cannot prove a hypothesis, but may disprove some of the working hypotheses and thus refine the set of hypotheses the scientist entertains to design further experiments.

Autonomous experimentation is an emerging concept in experimentation [1,3]. Moving beyond the traditional design of experiments, autonomous experimentation applies closed-loop experimentation, which uses the results of experiments conducted so far to decide which experiment should be performed next. Autonomous experimentation requires the combination of software to determine the experimentation that is to be carried out, and hardware to provide the platform for conducting the experiments.

Experimentation is also guided by resource management. When resource availability is limited, as it often is in experimentation, there is a trade off between information gained and resources used. Therefore, algorithms that maximise the amount of information gained within available resources are of particular interest.

Component Based Design

Splitting the idea of an autonomous experimentation algorithm into separate components, such as hypotheses management, experiment generation, experiment selection and result interpretation, allows for individual development of core aspects of an autonomous experimentation algorithm. Different ideas for each component can be rapidly tested, by plugging in different implementations of each component into the experimentation system.

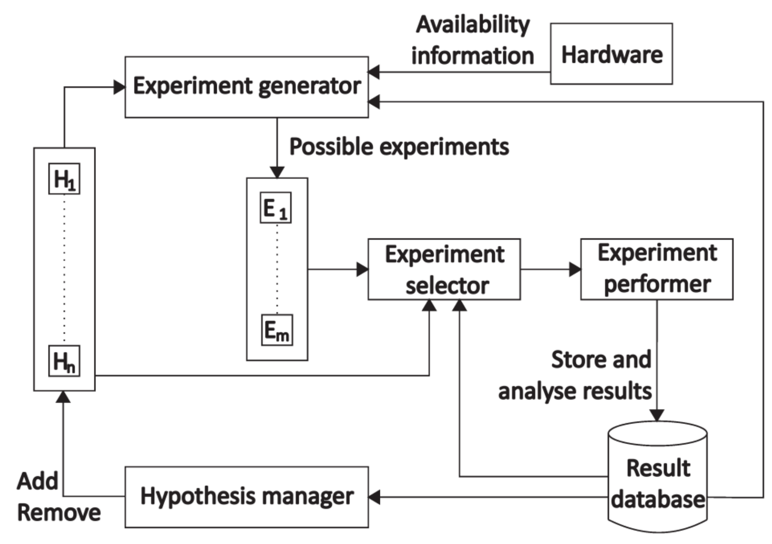

Figure showing component based design. $H_1 \dots H_n$ shows the working hypotheses. $E_1 \dots E_m$ shows the proposed experiments.

A Candidate Algorithm

Each experimentally observed result is used by the algorithm to generate a level of confidence that the result will be obtained across the experiment parameter space. Confidence shows the degree to which the hypothesis believes that the given result will be obtained when some experiment is performed, as shown in the figure below. The confidence for a result in an area of the experiment parameter space increases when supporting results are obtained and decreases when results that do not support it are obtained.

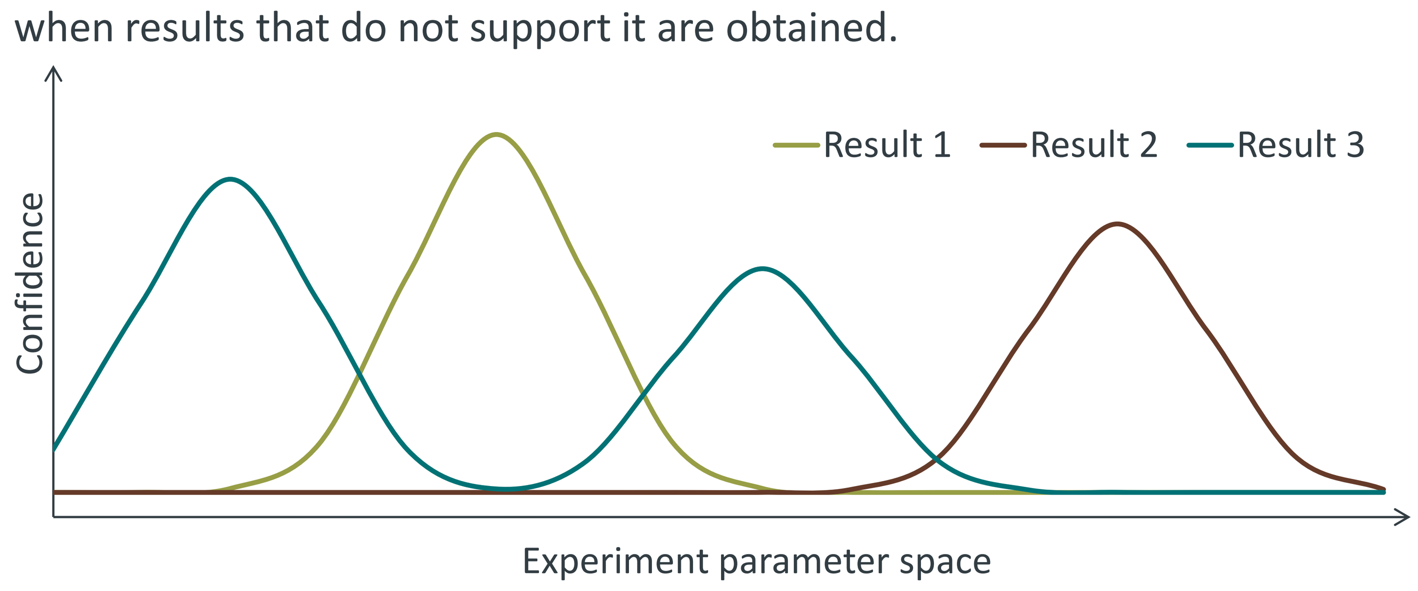

Figure showing the confidences that a particular result will be obtained by experiments performed with parameters from the parameter space.

This algorithm was tested on a simulated toy experimentation problem. In the problem, there is a discrete two-dimensional space, where each point in the space has a value assigned to it. In a real problem, this value would be some experimental result data. The space is segmented by a series of grids, with increasingly finer resolutions. Experiments simulated in this space can only observe the average value of the points within a particular grid square.

In the figures below, the values of each point in the parameter space are depicted as the colour of the pixels. After performing an initial exploration of the space using the lowest resolution grid, the algorithm determines for which areas of the space the confidences are low. In the figures below, these areas are blacked out. The system then continues to explore the space by focusing on the areas of the space where the confidences are low. Increasingly higher grid resolutions are used in the figure below (left to right).

Graphical illustration of candidate algorithm

Original | Initial pass | Second pass | Third pass | After five passes

References:
[1] D. Kulkarni and H. A. Simon. Experimentation in machine discovery. In J. Shrager and P. Langley, editors, Computational Models of Scientic Discovery and Theory Formation, pages 255-273. Morgan Kaufmann Publishers, San Mateo, CA, 1990.
[2] T. Matsumoto, H. Du, and J. S. Lindsey. A parallel simplex search method for use with an automated chemistry workstation. Chemometrics and Intelligent Laboratory Systems, 62:129-147, 2002.
[3] S. Muggleton and K.-P. Zauner. Artifical scientists. In S. Emmott and S. Rison, editors, Towards 2020 Science, pages 36-37. Microsoft Research, Cambridge, 2006.
[4] D. Gooding. Experiment and the Making of Meaning. Kluwer Academic Publishers, Dordrecht, 1990.