Knowledge-Based Systems 15 (2002) 349–354

Knowledge-Based SYSTEMS

www.elsevier.com/locate/knosys

Automated control of an actively compensated Langmuir probe system using simulated annealing

L. Nolle[a,*], A. Goodyear[a], A.A. Hopgood[b], P.D. Picton[c], N.St.J. Braithwaite[a]

[a]Oxford Research Unit, The Open University, Berkley Road, Oxford OX1 5HR, UK
[b]Department of Computing, The Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK
[c]School of Technology and Design, University College Northampton, Northampton NN2 6JD, UK

Abstract

A simulated annealing (SA) method has been developed to deduce 14 Fourier terms in a radio frequency waveform for active compensation of a Langmuir probe system. The active compensation system uses seven harmonics to generate a required waveform. Therefore, 14 heavily interacting continuous parameters need to be tuned before measurements can be taken. Because of the magnitude of the resulting search space, it is virtually impossible to test all possible solutions within an acceptable time. An automated control system employing SA has been developed for online tuning of the waveform. This control system has been shown to find better solutions in less time than skilled human operators. The results are also more reproducible and hence more reliable. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Simulated annealing; Optimisation; Langmuir probe system

1. Problem description

This article concerns the use of simulated annealing (SA) to control a diagnostic measurement system for a range of important industrial processes based on plasma technology. The plasmas in question are produced by the passage of electricity through a gas at low pressure, resulting in a partially ionised medium comprising electrons, ions, radicals and neutrals, according to which gases are involved.

The semiconductor industry has been a major exploiter of plasma technology. It uses plasma processes to etch and deposit the different thin layers of semiconductor, insulator and metal that make up integrated circuits. The demand for the ability to maintain uniformity and yield while scaling up for larger-area processing has promoted a considerable research effort in real-time assessments and control of the uniformity of technological plasmas.

Under the prevailing conditions, the plasmas used in materials processing are not in thermal equilibrium so considerable energy (typically equivalent to tens of thousands of Kelvin in temperature) resides in the electrons. Through interactions with these electrons, material can be broken into its constituent atoms. Thus gases and solids can be ‘activated’ electrically by means of a plasma. The heavier particles in plasmas tend to remain relatively cool so that the plasma medium facilitates the treatment of thermally sensitive surfaces and bulk materials. For a number of operational reasons, considerable use is made of radio frequency (RF) power supplies in the generation of technological plasmas. Basic electrical measurements on the distribution of charged particles in a plasma is a fundamental requirement for process developers and has potential as a routine control parameter in future.

The problem that is tackled in this article concerns the adaptation of a basic electrical measurement of uniformity that is much used in plasmas generated by direct current. When RF power is used to create the plasma, the electrical environment requires special attention. It transpires that one of the approaches, the one adopted here, though simple in principle, calls for a level of optimisation and control for which some kind of AI is essential.

1.1. Radio-frequency driven discharge plasmas

Under normal conditions, gases do not conduct electrically. Almost all electrons are bound to an atom or molecule. However, if electrons are introduced and given enough energy by an external energy source, like electro-magnetic fields, light, heat, etc. then they have the potential on collision with gas atoms or surfaces to release more electrons, which themselves may release other electrons. This resulting electrical breakdown is known as an avalanche effect. The so-formed ionised gas or plasma is now conducting.

* Corresponding author. Tel.: +44-1865-487-602; fax: +44-1865-326-322.

PII: S0950-7051(01)00171-X

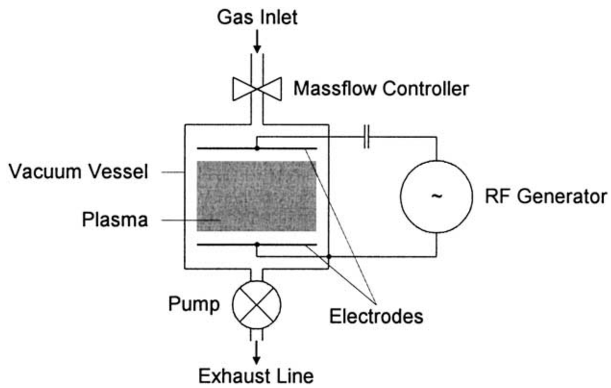

Fig. 1. Capacitively coupled RF-powered plasma.

In industrial RF powered plasmas, an RF generator is used as an external power-source, usually operating at 13.56 MHz. The use of RF rather than DC has developed for a number of reasons including efficiency and compatibility with systems in which direct electrical contact with the plasma is not feasible. The frequency of 13.56 MHz is assigned for industrial, non-telecommunications use. The RF is inductively or capacitively coupled into a constant gas flow through a vacuum vessel using electrodes (Fig. 1).

The electrons have only a fraction of the mass of the atoms, hence they can follow the electric field, while the ions respond only to slower variations in electrical structure. Electrons near the electrodes can escape. This results in electric fields, pointing from the plasma to the electrodes. These fields generate a flux of energetic ions, which can be applied continuously to a large area of workpiece, e.g. for etching or deposition.

1.2. Langmuir probes

Electrostatic or Langmuir probes were developed by Langmuir in 1924 and are one of the oldest diagnostic tools used to obtain information about properties of low pressure plasmas [1]. Such a probe consists of a small metallic electrode that is inserted into the plasma (Fig. 2). By applying a positive or negative DC potential to the probe, either an ion or an electron current can be drawn from the plasma, returning via a large conducting surface such as the walls of the vacuum vessel or an electrode. This current can be used to analyse the plasma, e.g. for the determination of the energy of electrons, electron particle density, etc.

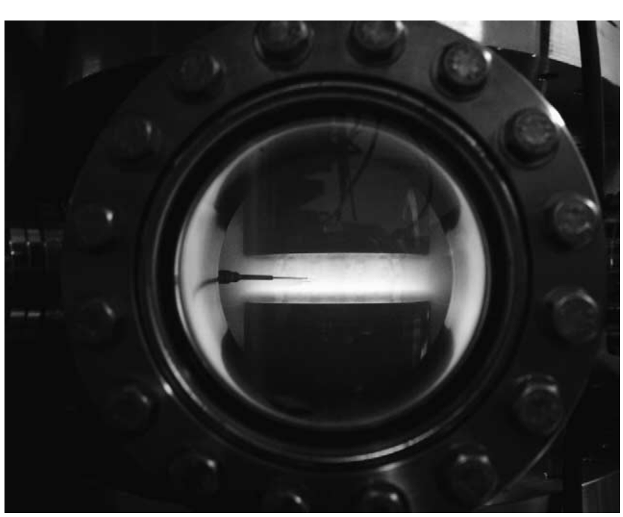

Fig. 2. Langmuir probe inserted into an Argon plasma.

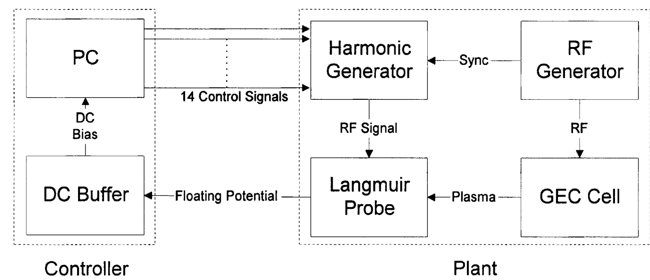

Fig. 3. Closed control loop for waveform tuning.

The region of space-charge (or sheath) that forms around a probe immersed in a plasma has a highly non-linear electrical characteristic. As a result, harmonic components of potential across this layer give rise to serious distortion of the probe's signal. In RF-generated plasmas this is a major issue as the excitation process necessarily leads to the space potential in the plasma having RF components.

1.3. Active compensation in RF-driven plasmas

In order to compensate for the time variation of RF potential difference between probe and plasma, the probe potential has to follow that of the exciting RF signal [2]. This can be achieved by superimposing a synchronous signal of appropriate amplitude and phase onto the probe tip. Plasmas are inherently non-linear and therefore they generate many harmonics of the exciting fundamental. As a consequence of that, the RF signal necessary for satisfactory compensation has not only to match in amplitude and in phase that of the exciting RF, it also has to match the waveform of the harmonics generated in the plasma.

Conveniently, the electrostatic probe and the plasma spontaneously generate a useful control signal. In the presence of a plasma, an isolated electrostatic probe adopts a 'floating potential', at which it draws zero current. The effect of inadequate compensation on a probe in RF plasma is to drive the DC potential of the probe less positive (or less negative). Thus, optimal tuning is identical with the probe adopting the most positive (or least negative) potential. The floating potential is also referred to here as a DC bias.

2. Automated control system

Previous work reported by Dyson et al. [3] involved the use of three harmonics for waveform synthesis. In this application, an additive synthesiser (harmonic box) with seven harmonics has been used to generate the appropriate waveform for a Langmuir probe system attached to a gaseous electronics conference (GEC) reference reactor [4]. Fig. 3 shows the schematics of the control system for waveform tuning. The 14 input parameters (seven amplitudes and seven phases) are heavily interacting due to the technical realisation of the synthesiser.

For example, the slightest departure from an ideal sinusoidal shape in one of the channels introduces harmonics itself. In practice, even after careful electronic design, it is found that there is a weak but significant coupling between the control of amplitude and phase. As a consequence of this, the number of points in the discrete search space has to be calculated as follows:

$$n = (2^b)^p \quad (1)$$

where n is number of points in search space, b resolution per channel in bits and p is the number of parameters to be optimised.

The D/A and A/D converters used in this project had a resolution of 12 bits and the dimensionality of the search space was 14. Hence, the search space consisted of $n \approx 3.7 \times 10^{50}$ search points. In this case, mapping out the entire search space would take approximately 10^{41} years. This was clearly not an option!

In order to tune the Langmuir system within an acceptable timescale, a suitable search algorithm had to be found. In selecting an appropriate algorithm, two requirements emerged. Firstly, the algorithm had to be capable of dealing with a high degree of epistatis (the degree of parameter interaction). Secondly, because of the absence of auxiliary information about the plasma, like its transfer function, etc. the prospective algorithm had to be a black box search algorithm. Previous research has shown that both requirements are met by SA [5]. Therefore, this algorithm was selected.

2.1. Simulated annealing

This general optimisation method was first introduced by Kirkpatrick et al. [6], based on the work of Metropolis et al. [7]. It simulates the softening process ('annealing') of metal. The metal is heated-up to a temperature near its melting point and then slowly cooled-down. This allows the particles to move towards an optimum energy state, with a more uniform crystalline structure. The process therefore permits some control over the microstructure.

SA is a variation of the hill climbing algorithm. Both start off from a randomly selected point within the search space. The difference between them is that if the fitness of a trial solution is less than the fitness of the current one, the trial solution is not automatically rejected, as in hill climbing. Instead it becomes the current solution with a certain transition probability $p(T)$, which depends on the difference in fitness and on the temperature T. Here, 'temperature' is an

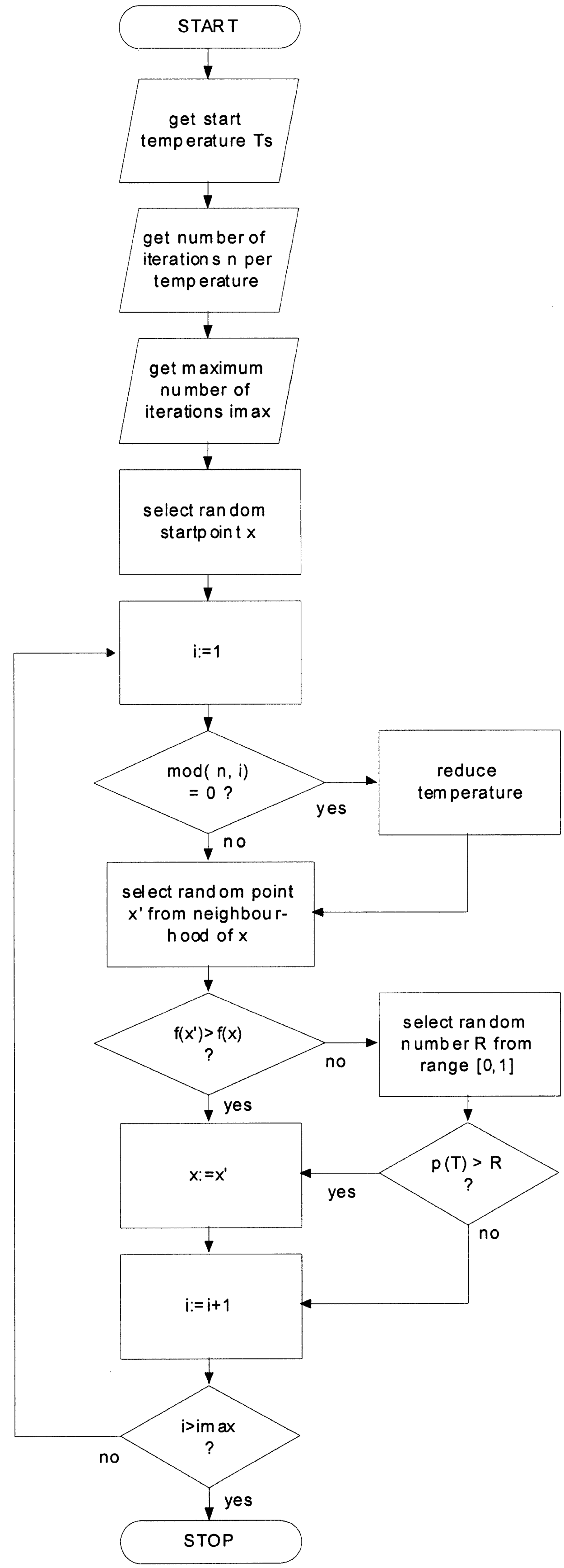

Fig. 4. Flow-chart of the basic SA algorithm.

abstract control parameter for the algorithm rather than a real physical measure. In this research, the transition probability $p(T)$ for a given temperature and a given difference in fitness ΔF is determined as follows:

$$p(T) = \frac{1}{1 + e^{\Delta F/T}} \quad (2)$$

where T is the temperature: control parameter for cooling schedule, $p(T)$ the probability of transition for temperature T and ΔF is the difference in fitness between current candidate solution and trial solution.

The algorithm starts with a high temperature, which then has to be reduced subsequently in steps.

$$T_{n+1} = \alpha T_n \quad (3)$$

where T_n is temperature at step n and α is the cooling coefficient.

On each step, the temperature must be held constant for an appropriate period of time (i.e. number of iterations) in order to allow the algorithm to settle in a 'thermal equilibrium', i.e. in a balanced state. If this time is too short, the algorithm is likely to converge against a local minimum. The combination of temperature steps and cooling times is known as the 'annealing schedule', which is usually selected empirically. Fig. 4 shows the flowchart of the basic SA algorithm.

3. The implementation of the control system

The system was set up at the laboratories at the Oxford Research Unit of the Open University. Fig. 5 shows the Langmuir probe system attached to the GEC reference reactor used during the development.

3.1. The development of the XWOS control software

Before the X windows waveform optimisation system (XWOS) control software was developed, the following requirements were identified:

- the optimisation should take place within reasonable time,
- the search results (fitness) over time should be plotted on-line on screen in order to allow a judgement of the quality of the result,
- the operator should be able to select values for the SA parameters,
- the operator should have the opportunity to set any of the 14 parameters manually,
- the operator should have the opportunity to fine-tune the settings found by the automated system,
- the DC bias (fitness parameter) had to be monitored.

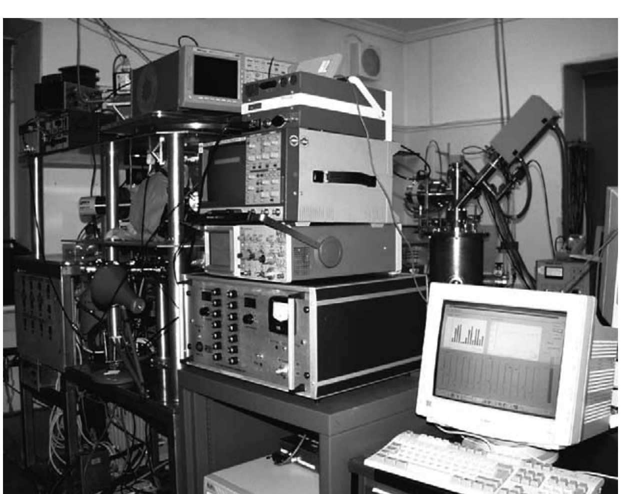

Fig. 5. Experiment set-up.

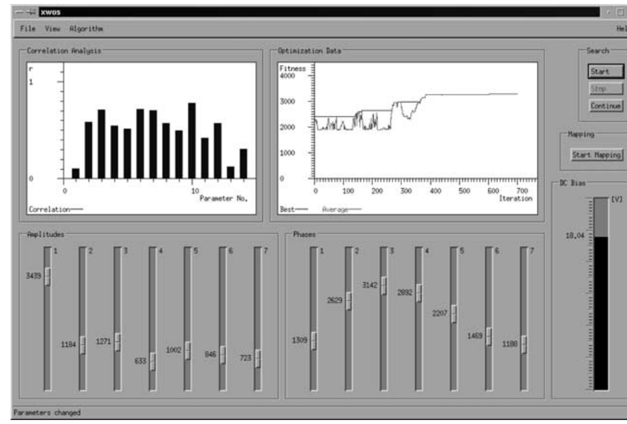

Fig. 6. XWOS main window.

The control software (Fig. 6) was developed in C++ on a 500 MHz Pentium III PC running the Linux 2.2 operating system. The graphical user interface was coded using X-Windows and OSF/Motif. XWOS allows the operator to start a search from the main window. During the search, the fitness over time (i.e. iterations) is plotted in real-time. After the algorithm has been stopped (either by reaching the maximum number of iterations or by the operator), the inputs for the waveform synthesiser, as well as the corresponding sliders on the main screen, are set to the best solution found during the search. The DC bias gained is displayed numerically and on an 'analogue' meter. The operator may then change the settings determined by the algorithm by hand, while observing the DC bias on the meter.

After a search, the history data of the run is used to perform a correlation analysis between the input parameters and the fitness function. The correlation coefficients are graphically displayed in a separate window and may be used as an indicator of the sensitivity of each channel for the given plasma parameters. The search parameters can be manipulated by the operator using the search parameter dialog box (Fig. 7). Here, the operator can also adapt the software to the plasma equipment by selecting an

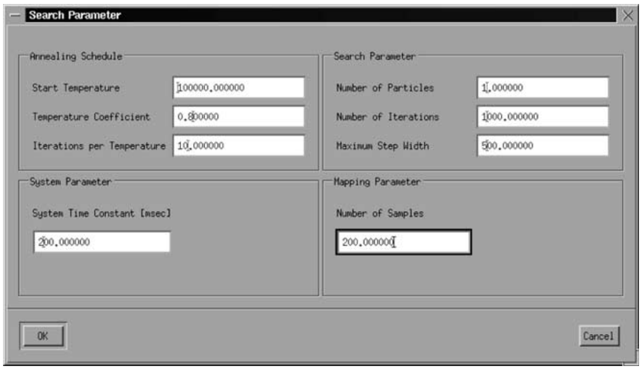

Fig. 7. Search parameter dialog box.

appropriate time constant for the plasma system. XWOS also provides the facility for random mapping of the search space for a more meaningful correlation analysis.

3.2. Choosing SA parameters

A sigmoid probability function (Eq. (2)) has been used in conjunction with the standard cooling function (Eq. (3)). The cooling coefficient α has been chosen to be 0.8. The maximum number of iterations was 1000, while the number of iterations per temperature was 20. The maximum step width, i.e. the maximum difference between the current solution and a new trial solution, has been set to 500. The selection of a suitable value for the start temperature T_S is most important for the success of SA. If it is not sufficiently large, the system starts in an already 'frozen' state. Hence, it is more likely to get stuck in a local optimum. If it is too large, the algorithm performs a random walk rather than a random guided search. Previous research has shown that a value of 0.5 for the initial transition probability $p(T_S)$ is a good compromise for a small ΔF, decaying towards zero for larger values of ΔF [8]. Therefore, T_S has been calculated as follows:

$$p(T_S) \approx 0.5$$

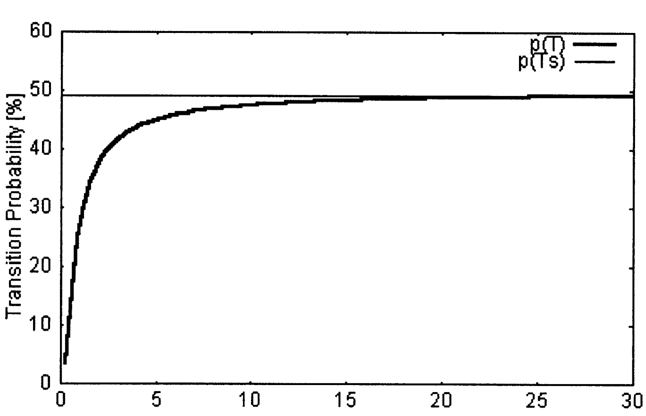

Fig. 8. Transition probability versus temperature.

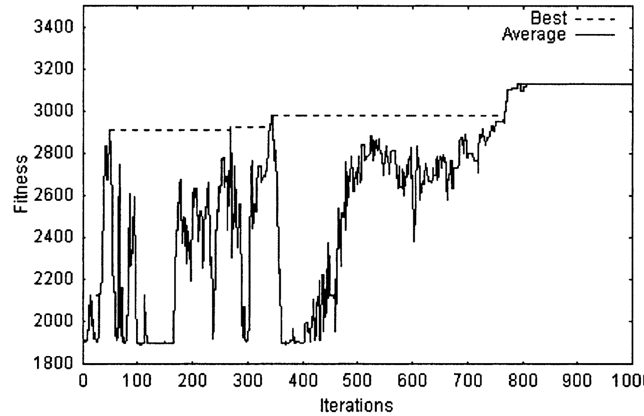

Fig. 9. Typical run of the SA algorithm.

Using Eq. (2)

$$\frac{1}{1 + e^{\Delta F/T_S}} = 0.49$$

Results in

$$T_S \approx 25\Delta F \tag{4}$$

Fig. 8 shows the graphical representation of the normalised probability function for $\Delta F = 1$ together with $p(T_S) = 49\%$. It can be seen that Eq. (4) can always be used to estimate the required start temperature T_S for expected values of ΔF if Eq. (2) is chosen to be the probability function.

4. System performance

Argon gas is used at a pressure of 100 mTorr. The plasma is usually produced using a power output from the RF generator of 50 W. Fig. 9 shows a typical run of the algorithm. The dotted line represents the best solution found so far, the thick line represents the fitness on each iteration. It can be observed that the algorithm initially explores the search space, while later on (after approximately 400 iterations) it exploits the most promising region. Hence, the

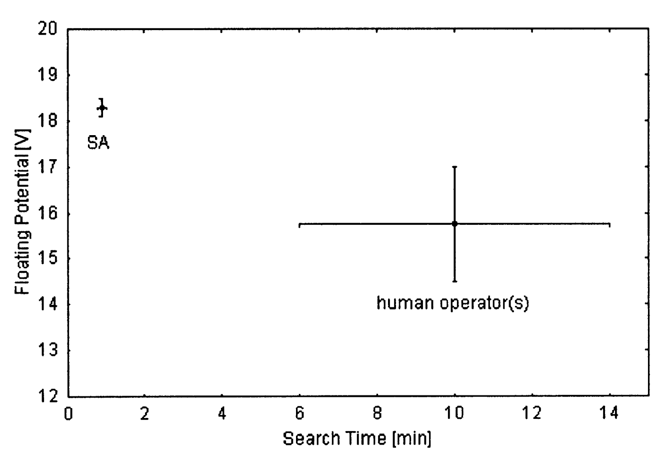

Fig. 10. Comparison between XWOS and human operators.

behaviour of the algorithm at the beginning of the search is similar to random walk, while towards the end it performs like ordinary hill climbing. The results gained during test trials by SA are compared with the results gained by different human operators in Fig. 10. The data points represent the average achieved DC bias (floating potential) plotted against the average time it took to find the solution. The error bars represent the deviations of these parameters. It can be seen that SA has always found better solutions than human operators, but has achieved these in significantly less time.

4.1. Discussion

While the quality (achieved fitness within consumed time) of the solutions found by SA shows only minor variations, the variation in quality for solutions found by human operators is quite dramatic. This is due to human aspects, like the operator's mood or experience. Hence, SA is more reliable in finding the global optimum and its solutions are more reliable. However, the existence of variations in fitness for SA indicate that the algorithm did not always find the global solution, even if it always came quite close to it.

It has been observed during the experiments that the selection of an appropriate step width is crucial to the performance of the SA algorithm. If the step width is too small, the algorithm has insufficient time to explore the whole search space, and is therefore likely to get stuck in a local optimum. If the step width is too large, the algorithm settles in the region of the global optimum, but if it is already near to the peak, new trial solutions are likely to 'fall off' the peak and hence miss the top of it.

Further work should include an examination of whether the step width can be adapted to the search space during the optimisation process and what effects this adaptation would have on the results achieved.

4.2. Conclusions

In this paper, a novel system for active compensation of a Langmuir probe system in an RF plasma has been presented. Seven harmonics have been used to synthesise the waveform of the potential as it appears in the plasma. Tuning of the 14 parameters (seven amplitudes and seven phases of the harmonics) by hand is not a straightforward task. The quality of the results depends on a number of human factors, like the operator's mood or experience. A software package has been developed to automate the tuning by SA. This type of optimisation algorithm was chosen because it offers two desired properties: it can cope with epistatis and noisy inputs and it is a black box optimisation method, i.e. it does not rely on auxiliary information about the system to be optimised.

The software has been used successfully for the tuning, and the results were compared with the results gained by human operators. SA clearly outperformed the human operators in both the fitness of the results and the time needed for the optimisation.

4.3. Further work

The next stage of this project will be the implementation of step-width adaptation. The effectiveness of such an adaptive system will be compared with the current system.

Acknowledgements

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) under grant reference GR/M71039/01. The authors wish to express their thanks to Fraser G Robertson, Faculty of Technology, the Open University, for assistance with the RF equipment, and to Pierre Barroy, Oxford Research Unit, the Open University, for his assistance in setting-up the experiment.

References

[1] J.D.S. Swift, M.J.R. Schwar, Electrical Probes for Plasma Diagnostics, Ilitte, London, 1970.

[2] N.M.P. Benjamin, N.St.J. Braithwaite, J.E. Allen, Self bias of an r.f. driven probe in an r.f. plasma, Mater. Res. Soc. Symp. Proc. 117 (1988) 275–280.

[3] A. Dyson, P. Bryant, J.E. Allen, Multiple harmonic compensation of Langmuir probes in rf discharges, Meas. Sci. Technol. 11 (2000) 554–559.

[4] M.A. Sobolewski, Electrical characterization of radio frequency discharges in the gaseous electronics conference reference cell, J. Vac. Sci. Technol. A—Vac. Surf. Films 10 (6) (1992) 3550–3562.

[5] L. Nolle, M. Walters, D.A. Armstrong, J.A. Ware, Optimum finisher mill set-up in the hot rolling of wide steel strip using simulated annealing, Proceedings of the Fifth International Mendel Conference, Brno, 1999, pp. 100–105.

[6] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.

[7] A. Metropolis, W. Rosenbluth, M.N. Rosenbluth, H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092.

[8] L. Nolle, D.A. Armstrong, A.A. Hopgood, J.A. Ware, Optimum work roll profile selection in the hot rolling of wide steel strip using computational intelligence, Lecture Notes in Computer Science, vol. 1625, Springer, Berlin, 1999, pp. 435–452.