CONF-951043--1

UCRL-JC-121676
PREPRINT

# Neural Manufacturing
# A Novel Concept for Processing Modeling, Monitoring and Control

Benjamin Law
Chi Yung Fu
Loren Petrich

This paper was prepared for submittal to the
SPIE Micro Electronic Manufacturing '95 Symposium
Austin, Texas
October 23 -24, 1995

**October 1995**

Lawrence Livermore National Laboratory



DLC

## DISCLAIMER

## Neural manufacturing -
## A novel concept for processing modeling, monitoring, and control

Chi Yung Fu, Loren Petrich, and Benjamin Law

Lawrence Livermore National Laboratory
University of California
P. O. Box 808, L-271
Livermore, CA 94550
Email: fu1@llnl.gov

### ABSTRACT

Semiconductor fabrication lines have become extremely-costly, and achieving a good return from such a high capital investment requires efficient utilization of these expensive facilities. It is highly desirable to shorten processing development time, increase fabrication yield, enhance flexibility, improve quality, and minimize downtime. We propose that these ends can be achieved by applying recent advances in the areas of artificial neural networks, fuzzy logic, machine learning, and genetic algorithms. We use the term neural manufacturing to describe such applications. This paper describes our use of artificial neural networks to improve the monitoring and control of semiconductor process.

Keyword List: Neural network, neural manufacturing, processing monitoring, fabrication control, adaptive process control, industrial productivity.

### 2. INTRODUCTION

We have identified a number of problems associated with traditional semiconductor manufacturing concepts. 1) Process and equipment modeling is still confined to the research laboratory, and is generally not available to front-line process engineers. 2) Existing forms of process control are not a closed-loop control of the entire fabrication step. They are instead a collection of disconnected, closed-loop control systems, concerned with peripheral parameters such as pressure and gas flow. 3) The process engineer has very few options available to rescue a "bad run," that is, a process that has deviated from specifications. 4) Some processing errors can be corrected by "reworking" or "processing feed-forward," but scheduling these reprocessing runs in the midst of fresh runs is a difficult logistical problem. 5) Extending the increasingly popular "cluster tool" concept to include metrology causes complications in tool design and integration. 6) It is customary to diagnose and treat equipment problems only after a failure has occurred. Prediction of failures in advance and undertake timely, pre-emptive repairs that would prevent costly impacts on manufacturing schedules have not been vigorously pursued. 7) Moving processes to different equipment often requires substantial parameter adjustments and downtime.

We believe that neural manufacturing concepts can solve many, perhaps all, of these problems. This will result in increased yield, greater flexibility, improved quality, and minimal downtime.

## 3. THE NEURAL MANUFACTURING APPROACH TO SEMICONDUCTOR PROCESSING

The semiconductor manufacturing process is extraordinarily complex, involving hundreds and thousands of individual steps. This complexity often prevents a complete scientific understanding of the mechanisms involved. As a result, we often use our human ability to recognize patterns and make inferences[1,2]. Our intuition and experience complement the precise number crunching calculations and the logical scientific deductions. It would therefore seem appropriate to call upon tools that mimic our human capabilities, tools such as artificial neural networks, fuzzy logic, genetic algorithms, and machine learning. Many of these tools are still in their infancy and are therefore limited in their capabilities, but even so they have significant advantages over their human counterparts. They will not, for example, make mistakes because of fatigue, and the "personal equation" is generally absent. We use the term neural manufacturing to describe the application of these human-imitating tools to the semiconductor manufacturing process. Research work by us[3] and others[4] strongly suggests that the effective use of these tools can greatly improve the return from semiconductor manufacturing lines. We believe that the continued development of neural manufacturing will have a fundamental and profound impact on the semiconductor industry.

## 4. EQUIPMENT CONTROL AND PROCESS MONITORING

In semiconductor processing of the usual kind, the actual fabrication process (etching, deposition, etc.) is not directly monitored or controlled. In plasma processing, for example, "process control" is done indirectly through an empirical ensemble of sensor outputs, conceptually very much like the methods of 20 or even 30 years ago. We monitor and control certain traditional parameters such as gas flow, RF power, and chamber pressure, but not the actual etching or deposition process. This indirect control methodology may fail to sense important aspects of the process. For example, certain types of deposition may be extremely sensitive to the presence of oxygen. The traditional methodology can sense only the oxygen flowing through the mass flow controller. However, oxygen resulting from water breakdown or outgassing from the chamber walls will not be detected. Thus we have only the illusion of process control. More importantly, the critical events at the wafer surface are not monitored and thus not readily under control.

It is obvious that traditional methods of process control ignore the vital reaction zone, i.e., the critical plasma region next to the wafers where fabrication actually takes place. Our neural manufacturing concept improves on the traditional methodology in two crucial respects:

1) The monitoring and control is done by an artificial neural network that can "learn" its complex task in somewhat similar manner as a human learns, and

2) We establish a continuous monitoring and control of the actual reaction zone through optical emission spectra and other appropriate data inputs.

## 5. IMPLEMENTATION OF NEURAL NETWORK PLASMA PROCESS CONTROL

In order to continuously monitor the reaction zone, which is right above the wafers, we choose a diode-array optical spectrometer[5] to provide the input signals to describe the reaction dynamics and kinetics. We have chosen an optical spectrometer for a number of reasons. First, the sensing is done externally and thus will not perturb the process itself. It is highly desirable to decouple the sensing from the processing equipment itself, if possible, because otherwise, if we move the process into different pieces of equipment, then the new processing equipment will have to be modified to accommodate the sensors internally. Also, the internal inclusion of the sensors into the processing equipment itself will generally lessen the degree of freedom for process development, especially when one chooses to do a totally different process in the same equipment. Second, the sensing itself is very fast and thus lends itself well to real-time monitoring and control. Third, optical spectrometers are generally readily available in the semiconductor industry. Fourth, unlike the residual gas analyzer which usually needs to have its head permanently connected to the equipment, the optical spectrometer is relatively portable, and thus minimizes the cost if sharing is possible. And finally, current end-point detection usually relies on optical sensing, and thus one does not have to have another set of instrumentation for end-point detection. In addition, more accurate end point detection can be achieved using a neural network based algorithm and thus can then be readily integrated with the our setup for monitoring and control. Optical spectrometer is not the only choice. Other instrumentation can also be used if it satisfies the key advantages mentioned above.

So, for a plasma-based processing system, plasma information is then obtained in the form of spectra from a diode-array optical spectrometer, which becomes the inputs for our neural-network system; the outputs of the network will represent information corresponding to the processing inputs for the system, such as gas flow, RF power, and pressure. With this *in-situ* real-time information pertaining to the plasma above the wafers, adaptive process control is then achieved by using this information to have the slave controllers make corresponding corrections (Figure 1). Correction signals will also then be tracked in the maintenance database to predict failures of the components other than the catastrophical ones and report to the maintenance engineers for preventive repairs at a time that will minimize interrupting the flow of the processes and the wafers.

We have chosen an oxide reactive-ion-etching (RIE) process to demonstrate the capabilities and performance of our neural network plasma control methodology. Experiments were performed in a batch-type, parallel-plate, processing reactor. The four processing input parameters were operational pressure, RF power, $CHF_3$ gas flow, and $H_2$ gas flow. Very high flow of $H_2$ gas were used, leading to very heavy polymerization, to push the processing into a highly unrealistic regime to check out our unorthodox approach in settings of extreme conditions. The spectrometer produced a 743-channel optical spectrum. Our center-cubic experimental design [6] minimized the number of experiments required to cover a very large, 4-dimensional operating space. A total of 30 experiments were performed; data from these experiments was used to train the neural networks. Additional experiments tested the performance of the system.

Raw data collected from the plasma emission are shown in Figure 2. The differences between the spectra are largely due to the attenuation of the optical emission by the polymer deposited onto the view port. After appropriate preprocessing, the resulting plasma spectra appear almost identical (Figure 3), even though the processing conditions are clearly different. The very heavily overlapping spectra coupled with the inherently noisy plasma environment present an extreme challenge for most signal processing. But artificial neural networks are known to perform well under these settings [1,2]. A neural network was successfully designed to distinguish the subtle differences between the inherently noisy plasma spectra and to identify the corresponding variations in processing conditions. The four outputs of the network represent the various gas flows, the pressure, and the power for the system corresponding to the plasma spectrum. The resulting network can accurately quantify the processing conditions for which it is trained with a maximum of 0.33% root mean square error (Table 1).

| **Parameter** | **RMS Deviation** |
|---|---|
| $H_2$ gas flow | 0.29% |
| $CHF_3$ gas flow | 0.19% |
| RF power | 0.33% |
| Chamber pressure | 0.25% |

Table 1. Summary of the neural-network results. The RMS deviation is between the neural-network results and the training data.

## 6. NEURAL NETWORKS ON A CHIP

To develop the full potential of the neural manufacturing concept, we believe that semiconductor fabrication processes must be sampled at many points over a large wafer area. This will require monitoring multiple sets of inputs from many different types of sensors located at strategic spots. The potentially enormous number of inputs gives us a formidable data-fusion problem. The neural systems will have to learn and "get smarter" as their training proceeds, feeding back more and more information to the process engineer. Such intelligent systems will be extremely complex, and may overwhelm the real-time capabilities of pure-software simulations. In other words, there are distinct limitations involved in simulating the function of neural networks, which are parallel-processing devices, through the use of software in serial-type computers. We believe that it is highly desirable to proceed to actual neural network hardware systems to ensure that we can handle the required data processing in real time in the future. This will require the availability of neural network hardware.

To this end, we have worked with the Naval Air Warfare Center at China Lake to design and fabricate our first neural-network, integrated-circuit chip. Such a chip is our first step in realizing the hardware systems that will be required for controlling complex and dynamic semiconductor processing systems. Our chip (Figure 4) contains 64 input-buffer amplifiers, 32 output neurons with individually-set gain and bias, and a fully-connected synaptic-array structure containing 16,384 synaptic resistors, which can function in either an excitatory or inhibitory mode. The present die size is 0.841 cm X 0.744 cm. Figure 5 shows a close-up view of the chip. The upper region is populated by the input buffer amplifiers. The bottom right-hand region is occupied by the neurons and the synapses are located in the bottom left-hand region. Given the layout of the chip is not dense and the technology is based on 2-micron geometry, we expect the size, and thus the cost, of this chip to shrink substantially as development proceeds. Such chips are designed to be cascaded to build more complicated systems. This chip is our first step toward realizing hardware systems for monitoring and controlling the complex and dynamic real-world semiconductor processing systems.

## 7. CONCLUSIONS

Within the limitations of this paper, we hope to convey the potential of the neural-manufacturing concept - a concept that integrates intelligence that mimics the human involvement in semiconductor processing through artificial neural networks, fuzzy logic and etc. to the area of manufacturing. This concept can be applied to any type of manufacturing that can provide appropriate inputs to appropriately-designed system such as artificial neural networks. For example, fluorescence or absorption spectra can provide input data for non-plasma-based processes. This neural manufacturing approach can improve yield by improving real-time process control; provide flexibility in the form of easily-reconfigurable manufacturing; and minimize downtime by predicting component failure and triggering pre-emptive repairs. The potential result is an increase in industrial productivity and reduction in cost.

There are still many problems to be overcome. For example, even though results obtained by ourselves and other researchers are very encouraging, we need further breakthroughs in the intelligent-system area. One major problem is that our neural network systems in their present form cannot receive and process higher-order abstract information from the human specialist. To solve this problem, we at LLNL has worked on a new type of intelligent system that is capable of including hypotheses and other such abstract knowledge provided by human experts. We call such a system a hypothesis-assisted intelligent system (HAIS). Such a system is fundamentally different from an artificial neural network or a fuzzy-logic system, and would act as a partner to a human researcher. The resulting partnership could explore and learn on a team basis. One form of this HAIS has already demonstrated very interesting results, along with a potentially much- higher degree of generalization than the neural network systems we describe here.

## 8. ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. DOE by Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.

## 9. REFERENCES

1. "DARPA Neural Network Study", Armed Forces Communications and Electronics Association International Press, Nov. 1988.
2. R. P. Lippmann, "An Introduction to Computing with Neural Nets," IEEE ASSP Magazine, April 1987.
3. C. Y. Fu, et al., "Application of a Neural Manufacturing Concept to Process Modeling, Monitoring and Control," Material Research Society Meeting, San Francisco, CA, April 17-21, 1995.
4. Y. L. Huang, et al., "Constructing a Reliable Neural Network Model for a Plasma Etching Process Using Limited Experimental Data" IEEE Transaction on Semiconductor Manufacturing, Vol. 7, No. 3, p. 333- 344, August 1994.
5. D. Malchow, "Characterization of Plasma Processes with Optical Emission Spectroscopy," SPIE Vol. 1392 Advanced Techniques for Integrated Circuit Processing, p. 498-505, October, 1990.
6. M. Jenkins, et. al., "The Modeling of Plasma Etching Processes Using Response Surface Methodology," Solid State Technology Journal, Vol. 29, no. 4, p. 175-182, April 1986.

Fu

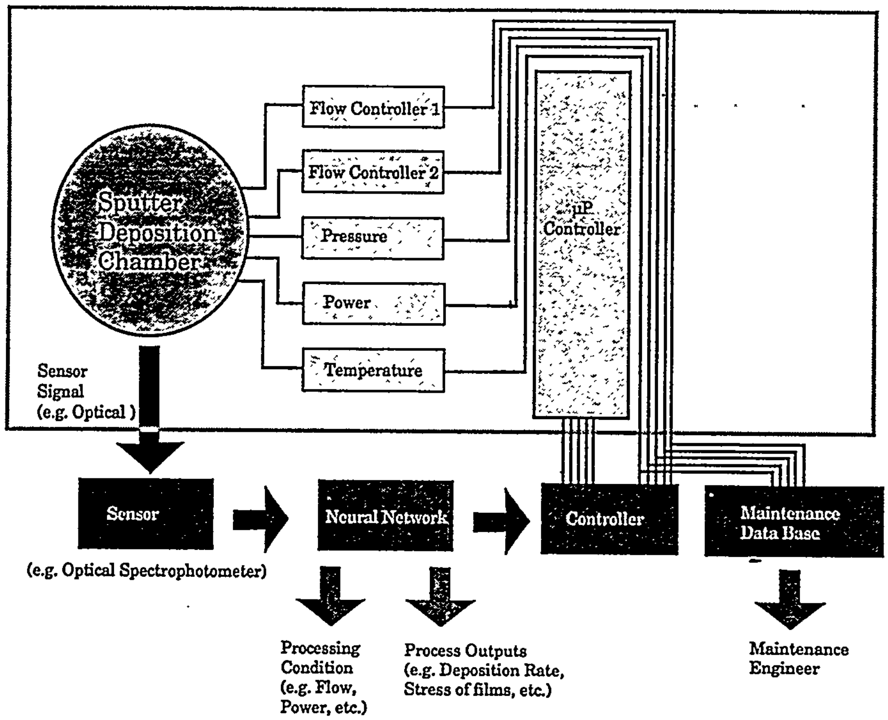


Figure 1 Schematic diagram of plasma monitoring and control using artificial neural networks. Information from the optical spectra is used to monitor and correct the process.

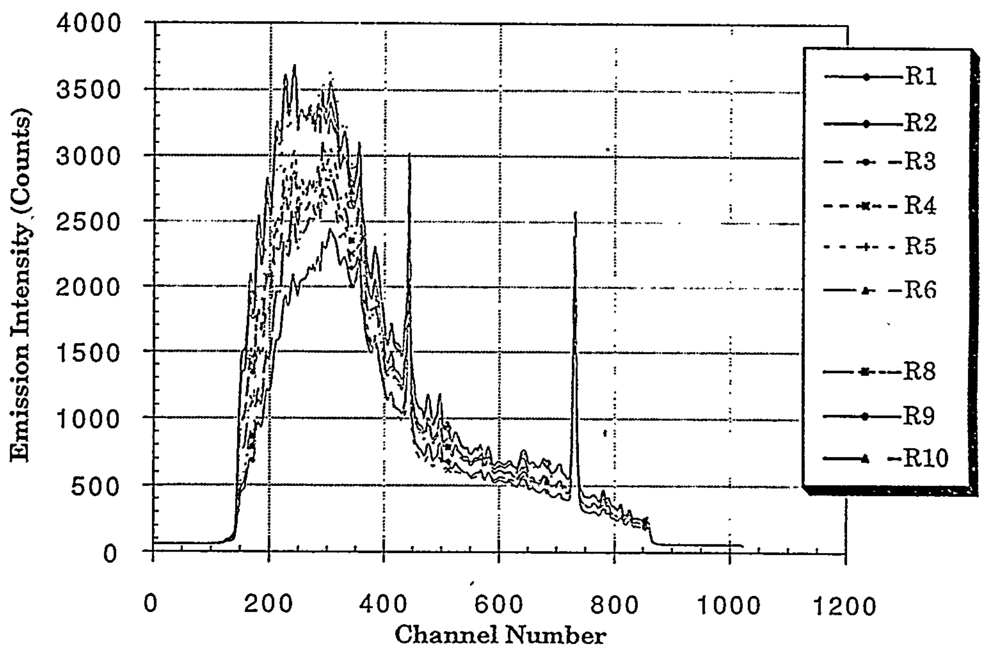


Figure 2 Optical spectra from some of the runs used to characterize the plasma. Heavy polymerization causes attenuation of the signals.

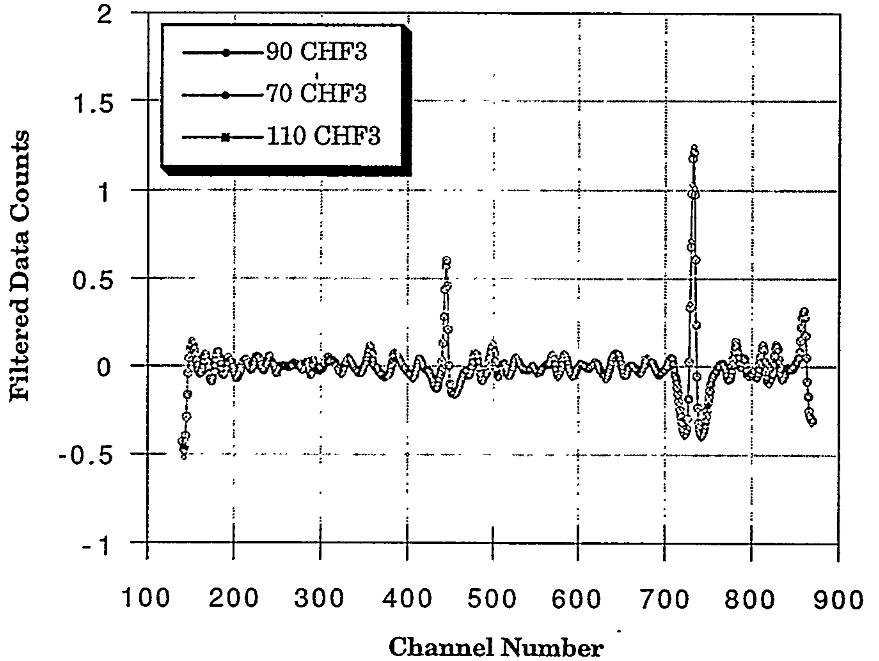


Figure 3 Comparison of preprocessed plasma spectra for various amounts of $CHF_3$ after preprocessing. Neural networks can distinguish the subtle differences among these spectra.

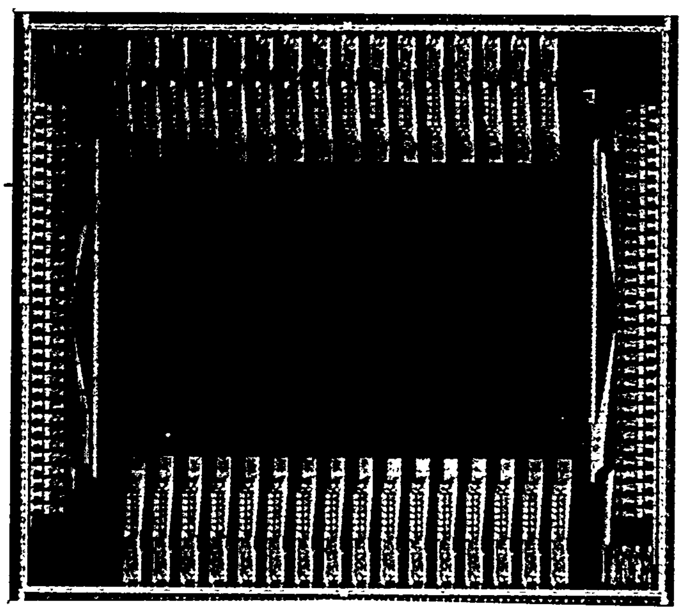

Figure 4 The LLNL/NAWC artificial-neural-network cascadable IC chip with 64 buffered input amplifiers and 32 output neurons (0.84 cm x 0.74 cm)

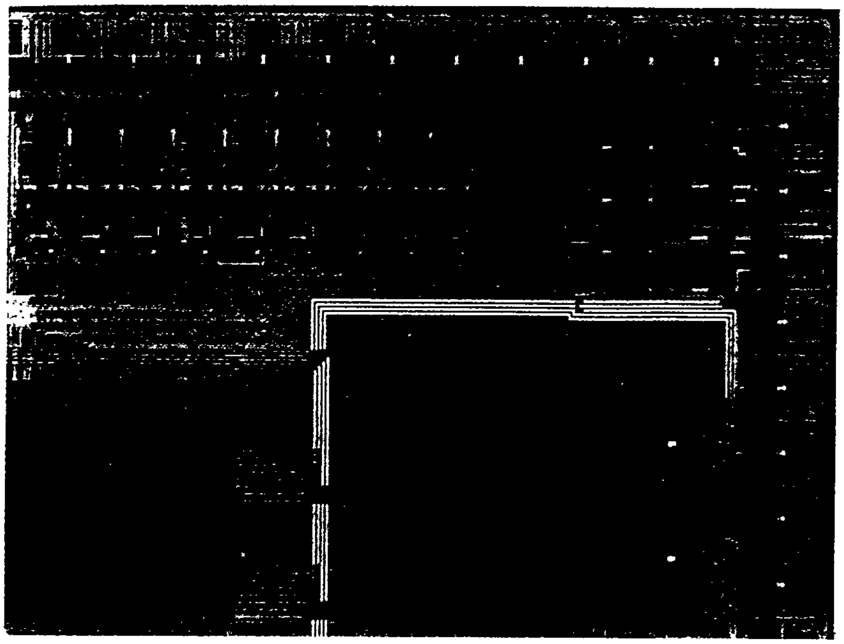

Figure 5 A close-up view of the LLNL/NAWC artificial-neural-network chip.

*Technical Information Department* • Lawrence Livermore National Laboratory
University of California • Livermore, California 94551