Real-Time, Heuristic-Based Control of Molecular Beam Epitaxy

O.D. Patterson, K.G. Eyink, and S. Cong

Real-time, heuristic-based control is appropriate for materials processes where accurate numerical models are not available. Frequently, the scientists working with a process base their decisions when controlling the process on a set of heuristics. This article describes the use of a rule base for real-time control of molecular beam epitaxy (MBE), a semiconductor thin-film growth process. As is often true in applying real-time control to materials processes, the most difficult task is to develop sensor technology capable of monitoring the material properties of interest. The use of ellipsometry, an optical technique, is described for the MBE process. The requirement to extract the material parameters from the ellipsometry data is computationally complex and time consuming. Development of algorithms to compress this manipulation into an acceptably short period for real-time control is discussed. Given these data, a rule base has been implemented to control the thickness and composition of thin films grown using MBE in real time. This research demonstrates that a rule base need not be complex to be effective.

Keywords

Ellipsometry, heuristic-based control, molecular beam epitaxy, real-time control

1. Introduction

INCREASING numbers of applications are using devices made from epitaxial III-V thin films such as heterojunction bipolar transistors (HBT) and high electron mobility transistors (HEMT).[1,2] For instance, the Department of Defense is sponsoring development of technology aimed at producing low cost, highly reliable monolithic integrated circuits under the Microwave and Millimeter Wave Monolithic Integrated Circuit Program (MIMIC).[3] Much of the current technology is based on ion implantation into bulk GaAs wafers, but an increasing number use epitaxial film-based HBTs and HEMTs. Other areas where the requirements for production level quantities of III-V semiconductor epitaxial films are increasing include digital high-speed circuits and electro-optic devices such as focal plane arrays.[3]

High-end epitaxial III-V thin-film production is dominated by molecular beam epitaxy (MBE).[2] This is the only technology capable of providing the quality material needed for many types of devices.[2,4] Unfortunately, the initial expense of using MBE may be prohibitive, with equipment acquisition costs of approximately one million dollars. Also, this process is comparatively slow, as typically only one substrate is deposited at a time at a rate of 1 μm/h. Finally, capable MBE operators command a high salary.

These economic factors limit MBE to high-end applications that cannot be satisfied with lesser material. These applications represent a small percentage of the total semiconductor output of the United States. Alternate processes such as chemical vapor deposition (CVD) and liquid phase epitaxy (LPE) provide the majority of thin-film material being used.

Reducing the cost of MBE will allow more applications to use materials grown from the MBE process, resulting in improved performance. A principal limitation of MBE is yield. Cooper assumed 80% as a typical production yield in modeling the economics of MBE.[2] His analysis showed that improvement in epitaxial layer yield from 80% to nearer 100% would result in nearly a 50% reduction in epitaxial growth costs.[2]

Presently, MBE, like many semiconductor production processes, is operated using open-loop control.[5] Subprocesses, such as furnace control, use closed-loop control via PID controllers; however the thin-film parameters are not involved in the control. A comprehensive plan for improving the manufacturing capability of MBE has been developed at the USAF Wright Laboratory.[3] One component of this system is heuristic-based, real-time control.

The challenges in developing real-time, closed-loop control for MBE and other processes in general include development of nonintrusive sensor technology capable of monitoring the material parameters of importance and a heuristic model of the process. This model makes it possible to adjust the control inputs based on the sensor data to achieve the desired performance.

A promising optical sensing technique, ellipsometry, is being developed to sense thin-film parameters of importance including film thickness, stoichiometry, and surface temperature. The ellipsometer measures the polarization state of light reflected off the thin-film surface. These data must be manipulated to find the composition and thickness of the thin film. Unfortunately, this data manipulation can take a great deal of time, because the equations that must be solved are transcendental. Traditionally, these equations have been solved using an iterative search technique; however, this method is too computationally intensive for real-time control.[6] Urban proposed using an Associative Neural Network (ANN) to obtain initial estimates to the solution allowing the interactive search to converge more rapidly.[6] This paper reports several improvements that make this method more practical for real-time control. The accuracy of the ANN has been improved. The topology of the

O.D. Patterson and **K.G. Eyink**, Manufacturing Research, WL/MLIM, Wright-Patterson Air Force Base, OH 45433-6533; and **S. Cong,** Technology Assessment and Transfer, Inc., 133 Defense Highway #212, Annapolis, MD 21401.

network has been adjusted to enable better modeling of data points on the boundary of the domain of interest. The concept of a pseudosubstrate has been incorporated, and most importantly, more data points are being used to minimize the errors due to experimental uncertainties.

Given the data provided by ellipsometry, a small rule base can be used to control MBE and achieve significant improvements in the yield of the process. Rule-based control can be applied when an accurate numerical model of the system is not available, as in the case with MBE. Advantages of rule-based control over modern control include: (1) the knowledge-base is easy for the MBE operator to understand and modify; (2) feedback control is easy to implement; and (3) a precise numerical model of the system is not necessary.

The literature provides examples of rule-based control in a wide range of applications.[7,8] Because ellipsometry monitors the particular parameters of interest, MBE can be controlled with relatively few rules.

2. MBE Process

Molecular beam epitaxy is a high-precision technique for growing thin-film semiconductor crystals that was developed in the late 1960s at Bell Laboratories by Cho and Arthur.[9] In MBE, a substrate typically of 77-mm diameter is suspended in the center of a vacuum chamber, called the growth chamber, which has a base pressure of less than 10^{-10} torr. Up to eight small ovens, called Knudsen cells, adjoin the growth chamber. Each Knudsen cell consists of a crucible loaded with a particular element such as gallium (Ga), arsenic (As), or aluminum (Al) and a furnace that is used to heat the element to its vaporization temperature. The crucibles are screened from the growth chamber by shutters. When a shutter is opened, a beam of atoms or molecules from that Knudsen cell is emitted toward the wafer. The magnitude of the flux depends on the temperature of the material in the crucible. Generally, multiple shutters open concurrently. The proportion of the atoms/molecules that bond epitaxially to the substrate depends on a number of factors including the substrate temperature.

Molecular beam epitaxy provides precise control over the semiconductor material being grown and therefore produces devices of superior quality. Deposition of less than a single monolayer of thickness can be controlled.[4] Because growth occurs at a relatively low temperature (580 to 620 °C for GaAs), bulk diffusion is minimal, and doping profiles are not disturbed. In addition, the dislocation densities, interface abruptness, mobilities, and minority carrier lifetimes of MBE-grown films are generally equal or superior to those grown by other state-of-the-art epitaxial techniques.[10]

A variety of factors contribute to the quality of the material grown by the MBE process including:

- Accuracy of material layer thicknesses
- Accuracy of alloy concentration
- Accuracy of dopant concentration
- Impurity levels
- Smoothness of material layer interfaces
- Crystalline defects

The nature of the MBE process makes it difficult to monitor. The process is very sensitive to impurities. Therefore, anything mounted inside the growth chamber must not outgas. Electron beam techniques should not be used to monitor the process because they add electron energy to the crystal, thus causing defects. In addition, the hot electron beam filament is a source of impurities during growth. Sensors cannot block the flux beams, as that would influence the flux distribution over the wafer surface. Given these limitations, an externally mounted sensor (ellipsometer) that uses photons to monitor the thin film was selected for development to control the MBE process.

Ellipsometry is capable of monitoring material layer thicknesses and alloy compositions. The impact of improving the accuracy of material layer thicknesses and alloy composition is difficult to predict because the level of achievement for these factors using current techniques is difficult to measure. One MBE researcher reported thickness variations of up to 10%.[11] This 10% thickness in a doped AlGaAs layer in a GaAs/AlGaAs depletion mode high electron mobility transistor (HEMT) resulted in a significantly altered threshold voltage, almost 1 V lower. With real-time monitoring of the thickness, accuracies of greater than ±1 monolayer are expected.

3. Ellipsometry

Ellipsometry is an optical technique that measures the change in polarization state that light exhibits upon reflection from a sample. This process includes the generation of light in a known polarization state, the interaction of this light with a sample, and the characterization of the polarization state of the exiting beam. The layout of an MBE machine equipped with an ellipsometry system is shown in Fig. 1. The results reported in this paper were obtained using a rotating analyzer ellipsometer. A helium/neon laser is linearly polarized and shone onto a sample at an angle of incidence of ~70°. The intensity variation of light over the revolution of the analyzer is measured at 5° intervals and averaged over ten analyzer revolutions. From these intensity measurements, ellipsometric angles Ψ and Δ are obtained every 5 s. Equation 1 defines Ψ and Δ, where $R_p(R_s)$ is the complex amplitude reflectivity for light polarized parallel (perpendicular) to the plane of incidence.

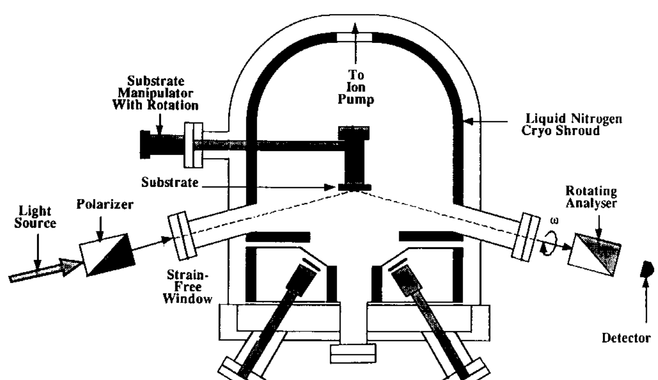

Fig. 1 Schematic of MBE machine outfitted with ellipsometry system.

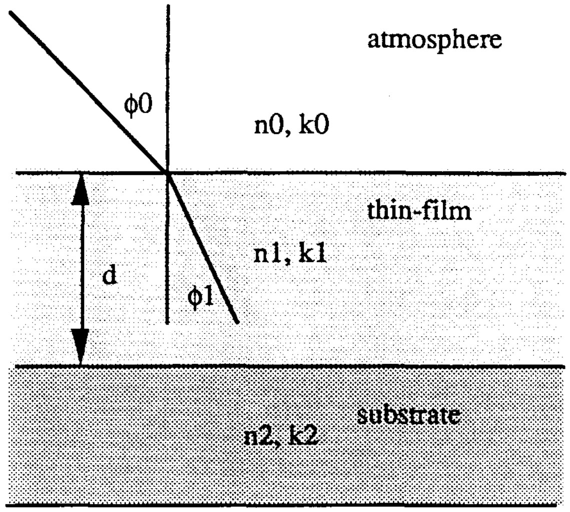

Fig. 2 Thin-film diagram.

The ellipsometry system can be set up in a variety of ways. One type of ellipsometry that has a clear advantage for real-time control is phase-modulated ellipsometry. Instead of using a rotating element, an electronically controlled phase modulator is used to modify the polarization state of the incident light beam. Hence, Ψ and Δ measurements may be extracted on the order of a few milliseconds. In addition, spectroscopic capabilities can be achieved by using a white light source. Each additional wavelength provides further information about the sample. The choice of wavelength is dependent on the material parameters to be monitored. Real-time spectroscopic capabilities may be realized with the use of a photodiode array.[12] For all these ellipsometry systems, inversion of the ellipsometry data is required for real-time control.

Given the complex index of refraction (n and k) and the thickness of the thin film (d), Ψ and Δ can be determined easily using Eq 1 through 8. The meaning of variables used in these equations is illustrated in Fig. 2. However, determining n, k, and d given Ψ and Δ, which is actually what needs to be done, requires the use of an iterative numerical search method such as the least-squares method. To find n, k, and d, three pieces of information are required. Therefore, at least two measurements must be made in quick succession. Assuming the stoichiometry has not changed during this brief interval (n and k remain constant), the unknowns are n, k, $d1$, and $d2$, whereas the measurements are $\Psi 1$, $\Delta 1$, $\Psi 2$, and $\Delta 2$. Given n, k, and the substrate temperature, the alloy composition can be found using reference data.

$$\tan(\Psi)e^{i\Delta} = \frac{R_p}{R_s} \quad [1]$$

$$R_p = \frac{r_{01p} + r_{12p}e^{-j2\beta}}{1 + r_{01p}r_{12p}e^{-j2\beta}} \quad [2]$$

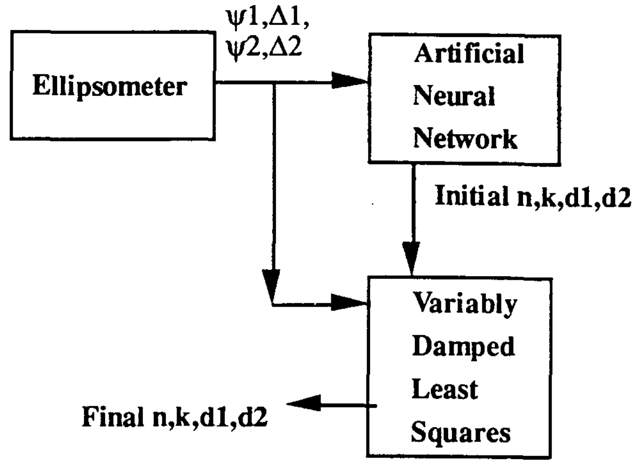

Fig. 3 Data inversion methodology.

$$R_s = \frac{r_{01s} + r_{12s}e^{-j2\beta}}{1 + r_{01s}r_{12s}e^{-j2\beta}} \quad [3]$$

$$\beta = 2\pi\frac{d_1}{\lambda}N_1\cos(\varphi_1) \quad [4]$$

$$N_i = n_i - jki \text{ (note here } j = \sqrt{-1}\text{)} \quad [5]$$

$$r_{ijp} = \frac{N_j\cos\varphi_i - N_i\cos\varphi_j}{N_j\cos\varphi_i + N_i\cos\varphi_j} \quad [6]$$

$$r_{ijs} = \frac{N_i\cos\varphi_i - N_j\cos\varphi_j}{N_i\cos\varphi_i + N_j\cos\varphi_j} \quad [7]$$

$$\varphi_i = \sin^{-1}\left(\frac{N_0}{N_i}\sin(\varphi_0)\right) \quad [8]$$

4. Ellipsometry Data Reduction

The two-step method proposed by Urban and shown in Fig. 3 allowed a least-squares algorithm to converge within ten iterations for the particular case reported.[6] Both components of this method of obtaining a solution to the transcendental equation problem may be improved. An ANN and least-squares algorithm called the Gauss Newton (Bard algorithm) were implemented in Think C™ from Symantec Corporation, Cupertino, CA on a Quadra 700 computer from Apple Computer Inc., to obtain the results discussed in this article.

A critical factor influencing the effectiveness of this algorithm is the amount of measurement error. For the ellipsometry system used for this article, the measurement error is one tenth of a degree. Use of additional Ψ and Δ data points by the least-

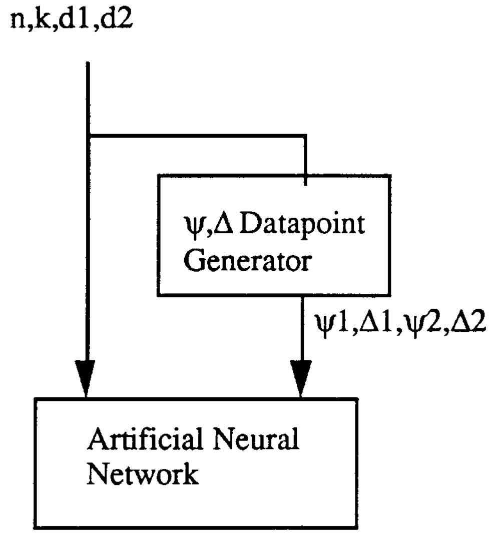

Fig. 4 Training configuration.

Table 1 RMS error versus number of measurements for d_{points} = 30 Å

Data Points	n	k	$d_{largest}$
2	0.0220	0.0084	1.80
3	0.0181	0.0068	1.72
4	0.0169	0.0065	1.59

Table 2 RMS error versus number of measurements for $d1$=15, $d2$=30, $d3$=45, and $d4$=60

Data Points	n	k	$d_{largest}$
2	0.0220	0.0084	1.80
3	0.0077	0.0025	1.04
4	0.0040	0.0010	0.71

squares algorithm is proposed as a method for reducing the effect of measurement error. The pair of experiments were run to demonstrate the merit of this idea. Random error of up to one tenth of a degree was added to sets of Ψ and Δ values generated using Eq 1 through 8. These values were then used to predict n and k. This experiment was repeated 1000 times for two, three, and four sets of Ψ and Δ values, and RMS errors were tabulated in Table 1. The desired values are n = 3.82, k = –0.2, and d_i = 30 Å, where i corresponds to the largest thickness. For two sets of Ψ and Δ, the RMS error of n is 0.022. Table 1 shows that, by using just two additional measurements, the accuracy of predictions for n and k can be improved by over 20%.

The ellipsometer used for the data in the reported research obtains data every 10 Å when the thin film is growing at 1 μm/h. With this limitation, the change in thickness between data points must be at least 10 Å. Table 2 compares the performance of using a two-, three-, and four-data point system where measurements are at 15-Å spacings. The results show that the decrease in measurement error due to the use of additional

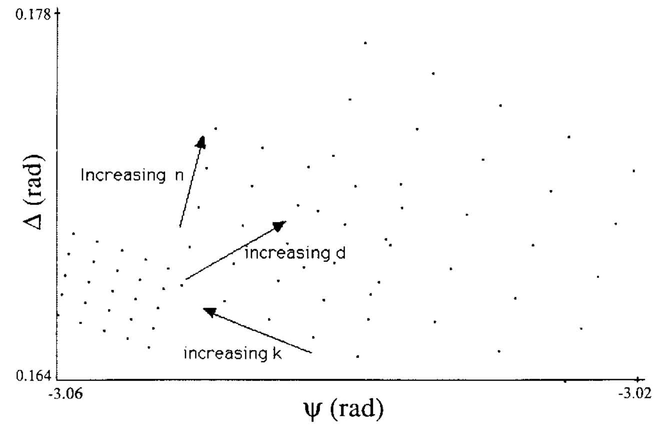

Fig. 5 Effect of increasing n, k, and d on Ψ and Δ.

measurements is even better when a consistent measurement rate is used. Improvements in n and k are over 80%.

The least-squares algorithm requires multiple iterations to arrive at a solution. The purpose of the ANN is to limit the number of iterations to ten iterations or even less. Ten iterations requires 0.73 s using the implemented least-squares algorithm, which is adequate for real-time control. Improved performance from the ANN would save even more time.

A common ANN, called a multilayer perceptron, with one hidden layer containing ten hidden nodes and multiple output nodes without nonlinearities, was eventually selected for the system. Urban noted that the data points lying near the outer limits of the training range were modeled much less accurately by the ANN than those in the middle of the range.[6] This problem can be avoided by removing the nonlinearity from the output neuron. The nonlinearity used is a sigmoid, as shown in Eq 9:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad [9]$$

To achieve an output of zero, a desired output given n, k, and d_i are normalized from zero to one, the input to this nonlinearity must approach negative infinity. Similarly, to achieve an output of one, the input to the sigmoid nonlinearity must approach positive infinity. These values are virtually impossible to obtain. Removing the nonlinearity from the output nodes enables values near zero and one to be obtained easily.

The multilayer perceptron was loaded with data from a data generator and trained for one million iterations (see Fig 4). Only two sets of Ψ and Δ were used in this case. All input values and output values were normalized. The RMS error in the ANNs estimate for n is 2.6% of the range of n. For k, it is 3.8%, and for d, it is 29.2%. These accuracies compare favorably to the accuracies reported by Urban, but should be improved further. A detailed analysis of the performance of the network shows that it trains to decompose variation in thickness into error components in n and k. The direction of increasing n, k, and d is shown in the plot in Fig. 5 where data points for n varying from 3.8 to 4.0, k varying from –0.2 to –0.5, and d at 10, 20, and 30 nm are plotted. For any Ψ and Δ, an infinite number of pos-

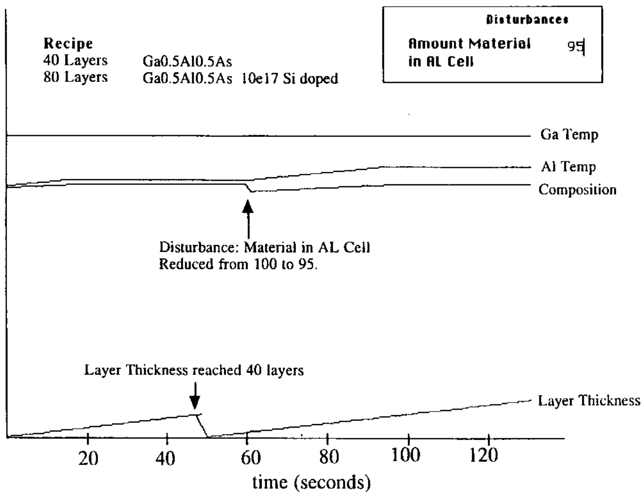

Fig. 6 Growth simulation using rule-based control.

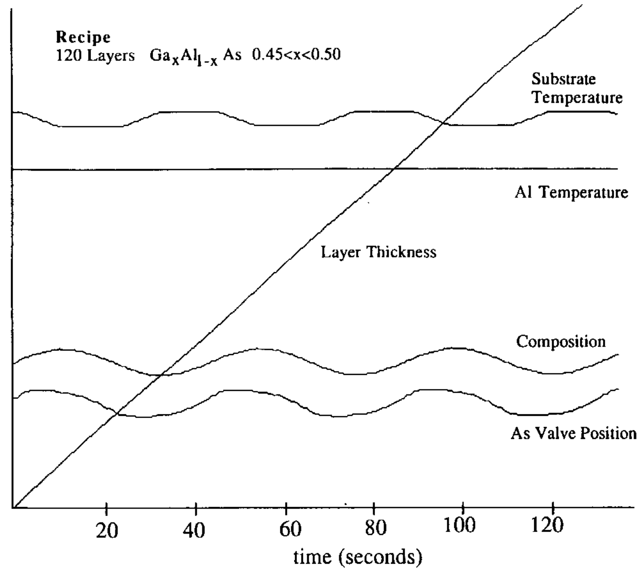

Fig. 7 Growth simulation for sinusoidal composition.

sible combinations of n, k, and d exist. Given the Ψ_1 and Δ_1 correspond to $n1$, k_1, $d1_1$, and $d2_1$, when the trained ANN is prompted with Ψ_2 and Δ_2 which correspond to n_1, $k1$, $d1_1+\Delta d1$ and $d2_1+\Delta d2$, it incorrectly predicts $n1+\Delta n$, $k1+\Delta k$, $d1_1$, and $d2_1$. Although the ANN is not able to predict the error component in the thickness value, its estimate is sufficient to bound the least-squares search to less than ten iterations.

The n and k found using Eq 1 through 8 are actually the average values over the entire thin film. If alloy composition is to be controlled, extraction of the instantaneous values of n and k in the currently deposited material is required. The Ψ and Δ measurements can be used to determine these values through the use of a pseudosubstrate. A pseudosubstrate is a fictitious substrate with n and k values giving the same optical characteristics as the existing stack of layers. After each Ψ and Δ calculation, the pseudosubstrate is updated to include the newly deposited materials. A new set of Ψ and Δ values is measured, and the next thickness is calculated from the surface of the new pseudosubstrate.

The use of a pseudosubstrate requires that the ANN contain several additional inputs, representing the changing complex reflectivity of the pseudosubstrate making the ANN more complex. On the other hand, the range of thickness over which the ANN must be trained is substantially reduced. Given accurate and current information about the composition and thickness of the thin film, the process can be controlled in real-time using a rule base.

5. Rule-Based Control

Conventional control of MBE is open loop through the use of a process plan. The length of time to grow each material layer and the set point temperatures of the Knudsen cells are predetermined based on characterization data from previous growth runs and calibration experiments run prior to the process run. Closed-loop control will reduce the effects of common disturbances and drift in the fluxes.

Rule-based control was selected rather than modern or conventional control for several reasons. First, an accurate analytical model of the process does not exist. Second, the requirements of the system do not demand performance beyond the capability of rule-based control. The dynamics of MBE, which are fast in comparison to some materials processes such as autoclave curing, are slow compared to many systems controlled by modern and conventional control techniques. Accurate regulation of the alloy concentration and layer thickness alone would be a significant improvement over state-of-the-art MBE processing. Varying alloy concentration is desirable for some applications and is possible with rule-based control. Third, the rule base is easy to understand and modify by the MBE operator. For instance, several control inputs exist to affect the alloy concentration; the operator can explore different algorithms for using the control inputs.

Figure 6 shows the results of a simulation of an MBE experiment using the rule-based control system. The vertical axis is dimensionless because the variation in the different parameters and not their absolute values is of importance. The "recipe" is given in the top left corner. Four signals are plotted: aluminum (Al) temperature, gallium (Ga) temperature, composition, and layer thickness. A process disturbance is invoked in this simulation; the amount of material in the aluminum Knudsen cell is suddenly changed from 100 to 95 units. The simulator determines that because there is less material in the aluminum cell, the aluminum flux will also be less, resulting in a decrease in the ratio of aluminum to all Group III atoms. The expert system is monitoring the instantaneous composition. Because the target composition is $Al_{0.5}Ga_{0.5}As$, the expert system recommends an increase in the aluminum temperature to restore the instantaneous composition. The layer thickness, which is the number of monolayers deposited in the current constituent layer, is reinitialized, and a new set of shutters is opened after each layer is completed. In Fig. 6, when layer thickness was reinitialized after 40 monolayers, the silicon (Si) shutter was opened.

Figure 7 shows a simulation to grow a material with a sinusoidal varying alloy composition. Two command inputs, the arsenic control valve and the substrate temperature, are used to control alloy composition. The arsenic control valve has a very fast response time, but its range is limited. The substrate temperature response time is slow, but it has an unlimited range.

This real-time, rule-based control system was also implemented in Symantec Think C™ and is a module of a comprehensive MBE control package being developed at the Air Force Wright Laboratory.[3]

6. Conclusions

A real-time, rule-based control system is being developed for MBE to effectively control material alloy composition and layer thickness. The knowledge base was described and demonstrated. In addition to a knowledge base, real-time control requires real-tim sensor data. Ellipsometry has been selected for application to MBE because its latency to the process and because of its ability to sense thin-film parameters including layer thicknesses and composition. Unfortunately, ellipsometry data must be reduced to provide this information. A two-step method for inverting ellipsometry data in real-time into thickness and composition data streams proposed in the literature has been improved. The first step is to obtain approximations of the solution to the inversion with an artificial neural network. Second, the Gauss Newton (Bard algorithm) is used to find the exact solution. The effectiveness of processing more than two data points simultaneously to reduce the effect of instrumentation inaccuracy was demonstrated. The concept of a pseudo-substrate to enable calculation of the instantaneous alloy composition rather than the average composition of the entire thin film was discussed. These improvements significantly contribute toward achieving real-time control of MBE thin-film thickness and alloy composition. Given that an increase in yield from 80 to 100% will result in a cost reduction of 50% in MBE material, real-time control of MBE will enable many more applications to use the high-quality material produced by MBE.

Acknowledgments

This research was supported by the Materials Directorate, Wright Laboratory, Wright-Patterson Air Force Base, Ohio. The authors would like to thank Mr. D.C. Moore for his technical advice at various stages of this research effort. Thanks are due also to Dr. W.M. Griffith and Dr. J.C. Malas, particularly for their help in the preparation of this manuscript.

References

1. K.D. Pedrotti, R.L. Pierson, Jr., C.W. Farley, and M.F. Chang, Monolithic Optical Integrated Receivers using GaAs Heterojunction Bipolar Transistors, *Microwave J.*, Vol 36(No. 5), May 1993, p 254-261
2. T.L. Cooper, MMIC Production Using MBE: Present and Future, *Microwave J.*, Vol 33(No. 6), June 1990, p 105-116
3. O.D. Patterson, J.J. Heyob, S.J. Adams, V. Hunt, P.H. Garrett, K.R. Currie, et al., "Progress Toward a Comprehensive Control System for Molecular Beam Epitaxy," WL-TR-92-4091, Aug 1992, Wright Laboratory, Wright Patterson Air Force Base, Dayton, OH
4. G. Bauer and G. Springholz, Molecular Beam Epitaxy—Aspects and Applications, *Vacuum*, Vol 43(No. 5-7), p 357-365
5. National Research Council Panel on Plasma Processing of Materials, *Plasma Processing of Materials: Scientific Opportunities and Technological Challenges,"* National Academy Press, 1991, p 30
6. F.K. Urban, III, D.C. Park, and M.F. Tabet, Development of Artificial Neural Networks for Real Time *in situ* Ellipsometry Data Reduction, *Thin Solid Films*, Vol 220(No. 20), 1992, p 247-253
7. D.A. Rowan, On-Line Expert Systems in Process Industries, *AI Expert Magazine*, Aug 1989
8. C.L. Dym, Expert Systems: A New Approach to Computer-Aided Engineering, *Engineering with Computers*, Vol 1, 1985, p 9-25
9. A.Y. Cho, Growth of III-V Semiconductors by Molecular Beam Epitaxy and Their Properties, *Thin Solid Films*, Vol 100, 1983, p 291-317
10. K. Ploog, Molecular Beam Epitaxy of III-V Compounds: Technology and Growth Process, *Ann. Rev. Mater. Sci.*, Vol 11, 1981, p 171-210
11. K.G. Eyink, private communication, Wright Patterson AFB, Ohio, June 1992
12. Y. Cong, R.W. Collin, and K. Vedam, Characterization of Ion Beam-Induced Surface Modification of Diamond Films by Real Time Spectroscopic Ellipsometry, *J. Vac. Sci. Tech.*, Vol A9(No. 3), 1991, p 1123-1128