MACHINE LEARNING FOR A DYNAMIC MANUFACTURING ENVIRONMENT

R. A. Perez and J. T. Lilkendey
University of South Florida

S. W. Koh
Harris Semiconductor

Introduction

Within the semiconductor wafer manufacturing process, tight quality control is of utmost importance. This is due to such factors as the highly competitive nature of the business [2] and the complex nature of the process itself. Therefore, it is imperative that processing problems are recognized and corrected as quickly as possible. To accomplish this, a Parametric Test facility exists where a critical quality check is performed on the wafers. This is accomplished by measuring a number of electrical parameters at existing test sites on the wafer. Each measurement must fall within an acceptable range of values for its associated parameter to pass the critical check. If test results indicate that a parameter is outside of its acceptable range, the wafer may fail the quality check and be scrapped. If this happens an expert must examine all the parametric data associated with that wafer and attempt to determine the reason for failure and where in the manufacturing process the problem may have occurred.

The expert's ability to diagnose these failures springs from both a knowledge of semiconductor physics and experience with the wafer manufacturing process. Since this combination of knowledge and experience is valuable and rare, it was decided that an expert system should be designed and implemented to take the place, or at least reduce the burden, of the human expert in the area of diagnosing wafer failures. As a result, the ADEPT expert system was created through the combined efforts of professors and graduate students in the Computer Science Department at the University of South Florida, in Tampa, Florida and process and development engineers at Harris Semiconductor, in Palm Bay, Florida. This expert system was deployed in August 1989 using a SUN 3/60 workstation.

Expert System Approach

ADEPT was designed with the following requirements:

1. It had to directly access the parametric wafer data stored in the pre-existing INGRES relational database.
2. It had to examine all wafers for grading and diagnosis. To keep up with production rates, each lot of wafers (approximately fifty), had to be diagnosed in approximately three minutes.
3. It had to incorporate all knowledge which the expert possessed about the diagnosis of failures, which included both procedural and heuristic knowledge.
4. Failed wafers could be diagnosed solely on the basis of the wafer data stored in the database, thus the system required no user interaction and could simply be triggered to run in the background when new data arrived in the database.
5. The users expected the diagnostic results to be stored in the database as well as having statistics and diagnostic summary reports generated.
6. The users were not concerned with how specific diagnoses were arrived at by the expert system.

Expert System Description

A diagram depicting the structure of the ADEPT expert system can be seen in Figure 1. Each module, except "Diagnose Wafers", is a separate program (coded in the C programming language) which runs in the background and stores results in the database. The "Diagnose Wafers" module is written in KEE (Knowledge Engineering Environment), a knowledge engineering development tool designed by Intellicorp^{TM}. It houses the knowledge base, which is composed of rules and frames, and a forward chaining inference engine. The input to the system is the parametric data which is transferred to the database from the Parametric Test facility through an ethernet. The database stores, on-line, one year of wafer data and diagnoses, and is approximately 300 Mbytes in size. Upon arriving in the database, the information flows among the different modules as depicted in Figure 2. Output from the system is stored directly in the database and accessed through associated Engineering Report Facilities.

The flow of information begins with the "Load Database" module storing the parametric data for each site on each wafer in the SITE_DATA table and recording the existence of new lots by storing their identity in the NEW_LOTS table. The "Check Against Limits" module utilizes the data stored in NEW_LOTS table to determine which lots need to be examined. It chooses the first lot present in the NEW_LOTS table and retrieves the parametric data associated with each site of each wafer in that lot from the SITE_DATA table. The

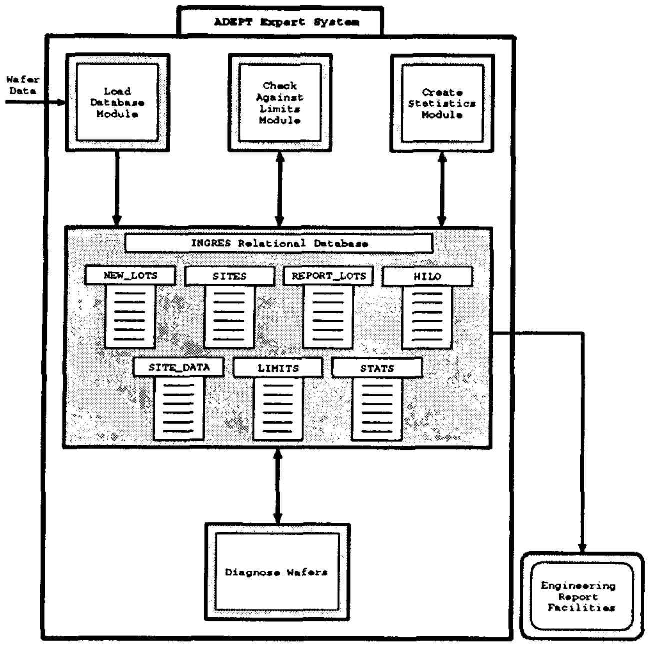

Figure 1: Structure Of The ADEPT Expert System.

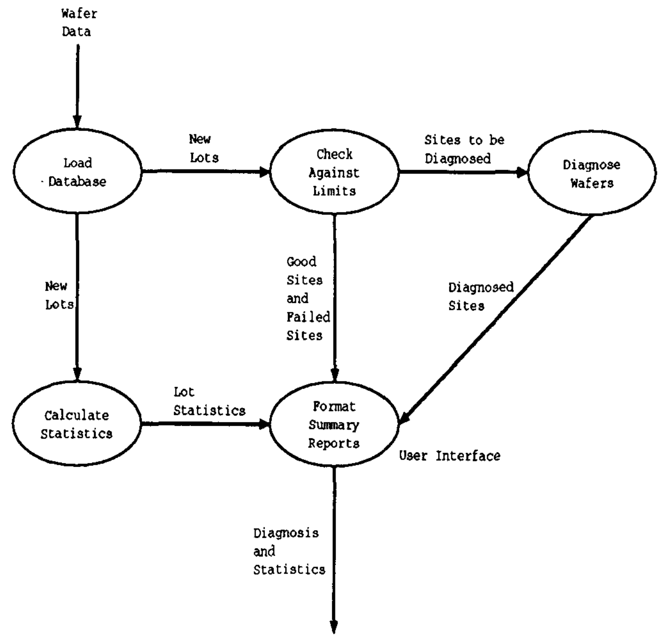

Figure 2: Information Flow in ADEPT.

data is compared with range limits which are stored in the LIMITS table and one of three decisions is made. Either (a) the measurements for the site are invalid, (b) it is a good site or (c) the site needs to be diagnosed. The module also converts all the parameter measurements into discrete equivalents which indicate if the value is low, normal or high as compared with the specification limits, diagnostic limits and invalid limits. The discrete values are then stored in the HILO table. Finally, the "Check Against Limits" module also handles the execution of the expert's procedural knowledge, mentioned in the previous section. Based on the relation of the parametric values to their associated limits, good and invalid sites are stored directly into the DIAGNOSIS table. The wafers whose sites requiring diagnosis are stored in the SITES table. When all wafers in a lot have been classified, the lot's identity is stored in the REPORT_LOTS table.

When a diagnosis is necessary, it is made by the "Diagnose Wafers" module. This module uses the discretized parametric data stored in the HILO table along with the expert rules in its knowledge base to diagnose the wafer sites stored in the SITES table. The resulting diagnosis is stored in the DIAGNOSIS table.

The "Create Statistics" module selects lots from the REPORT_LOTS table, and for each lot, calculates the lowest value, highest value, mean and standard deviation for each parameter. The result of these calculations are stored in the STATS table which is accessed by the Engineering Report Facilities.

Expert System Performance

The original ADEPT system was developed over the time period from August 1988 to August 1989 and required a total manpower effort of approximately one and a half person-years [6]. It was originally capable of diagnosing the wafers of a single flow, which is analogous to a single product line. It did so with a 90% degree of accuracy when compared to the expert, and this was judged to be acceptable.

From May to August of 1990, ADEPT was modified to allow it to handle an increased volume of data and two additional flows. This involved increasing system speed and efficiency and some additional standard knowledge acquisition. This work took approximately four person-months. Performance was determined to be at approximately 97%.

Like most expert systems, ADEPT was relatively static in nature. It was, however, being applied to a relatively dynamic environment. Although it was designed in a modular fashion and could be changed easily from a system standpoint, there was still the problem of changes to the knowledge base. As knowledge of a flow would change, there would be a need for continued knowledge acquisition and changes to the knowledge base. The addition of new flows to the knowledge base was an even greater problem, since the knowledge acquisition time was even greater. All of this suggests that it would have been necessary to constantly perform knowledge acquisition between the domain experts and knowledge engineers to insure system performance at an acceptable level over the desired range of products, in addition to making the appropriate changes to the expert system.

We estimated that each new flow would take an average of one person-month to implement. At the time there were approximately 70 flows passing through the facility [4] and we concluded that the amount of knowledge necessary to implement all 70 flows would have been impossible for the expert systems support group at Harris to manage. Assuming that the task could have been accomplished, it was estimated that it would have taken from two to three person-years to complete. In addition, as new knowledge was added it was associated very closely with a small number of processes. As a result, ADEPT would tend to perform very well when diagnosing wafers from these processes, but would be unable to deal with wafers from processes for which it had not been given rules. This problem began to grow as new processes were introduced and the knowledge engineers were unable to update the system at an acceptable rate. This is not to imply that ADEPT would wrongly diagnose problems which occurred on wafers from process with which it was unfamiliar. ADEPT had the ability to indicate if it was unable to diagnose a wafer. Although this is a good feature to have in any system, the problem was that, by the summer of 1991, ADEPT was unable to provide a diagnosis a large percentage of the time. In fact, in the period of about a month and a half, from April 4 1991 to May 16 1991, ADEPT was effectively saying "I don't know" about 34% of the time.

Therefore, it was decided to attempt to replace ADEPT with a system which would circumvent the previously mentioned difficulties and would be capable of operating for long periods of time without support from the expert systems group. The system chosen to replace ADEPT is described in the following section.

The ADEPT-II Classification System

ADEPT-II is composed of two distinct parts, the Dispositioner and ATI. The Dispositioner functions much like the "Diagnose Wafers" module in the previous expert system, in that it takes, as input, the parametric data associated with each wafer and produces a diagnosis. ATI, the ADEPT Training Interface, is an interface to the Dispositioner which allows users to alter the knowledge existing within it, thus affording the system a new dimension of user maintainability, which ADEPT previously lacked. We describe first the Dispositioner, and then the ATI.

A diagram depicting the structure of the ADEPT-II expert system can be seen in Figure 3 and it shows that only a few significant changes were made to the overall structure of the system. The "Diagnose Wafers" module which existed in ADEPT is gone. It has been replaced by the "Dispositioner". The Dispositioner is a module, written in C, which runs in the background, diagnosing (or dispositioning) wafers and storing the results in the database. The "Dispositioner" uses the GID3* algorithm [1] and the "Knowledge Base" to diagnose the wafers. The "Knowledge Base" is a decision tree which is generated by training GID3* on the main training

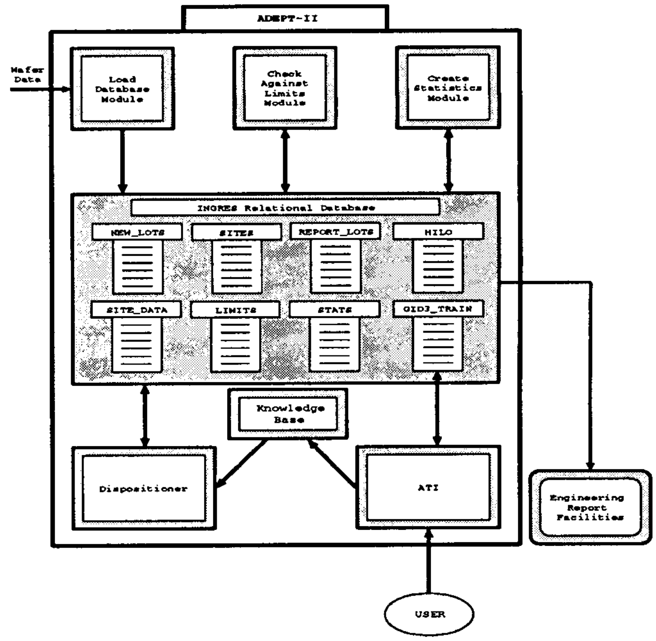

Figure 3: Structure Of The ADEPT-II Classifier.

set of examples, stored in the database. The name of the table which holds the training set of examples is "GID3_TRAIN".

A detailed description of GID3 is given in [1] and we include here only a very brief one. GID3 is a symbolic induction algorithm which, when given a set of previously classified examples, builds a decision tree that can correctly classify them. In our wafer diagnostic environment, each example consists of attributes and values of parametric measurements and the associated diagnosis. Normally, only a portion of the attributes are necessary for correct classification and these are the only ones included in the branches of the decision tree. If the examples are representative of the problems in the domain, then GID3 can diagnose wafers not included in the training set with accuracies of 95%. An evaluation of GID3 and several other machine learning tools in this environment can be found in [7].

When the dispositioner was originally implemented, it was capable of diagnosing the wafers from all flows which passed through the facility. The total number of possible flows was approximately 70 while the actual number of flows which the system "saw" from June 1 to August 1 of 1991 was 36. By the end of June, the decision tree used by the dispositioner, was created with a training set of just 180 classified examples. The performance at this time and at later dates is discussed in the following section. Additionally, it should be noted that the effort required to create both the dispositioner and the support system described below was approximately four and one half person-months, as opposed to the estimated two to three years it would have taken to modify the existing ADEPT system.

ATI, is a subsystem of ADEPT-II which allows users to handle the maintenance of the expert system's knowledge base with minimal aid from the knowledge engineers who created it. In ATI, user maintenance refers to allowing the user to change the knowledge contained in ADEPT-II through the introduction of new training examples. After the user recognizes that ADEPT-II is diagnosing wafers incorrectly, the user is able to identify those wafers by lot and ask the expert to diagnose them properly. At which time, ATI will automatically retrain using these new examples and the existing set of training data, contained in the "GID3_TRAIN" table. This process requires no further intervention by the user. After a new decision tree has been created, reflecting the new knowledge, the user is notified via e-mail of the systems completion and any problems which may have occurred. Thereafter the user may choose to commit the new knowledge to ADEPT-II, which will allow it to immediately use this knowledge in future diagnoses. Even after committing to use the new knowledge, the user may go back to using the previous decision tree. In fact, the system keeps four levels of complete backups, so the user can return to any of the last four decision trees. This type of backup system is important in a black box classifier, such as ADEPT-II, since the addition of new examples, especially wrongly classified ones, may have an adverse affect on the systems overall performance. As can be seen in Figure 3, the "ATI" module interacts only with the database, the knowledge base and the user, but not with the dispositioner itself.

This system was written in C and EQUEL, an embedded INGRES query language and it is presently running on a SUN 4 workstation.

Performance of ADEPT-II

After the initial deployment of the "Dispositioner", testing to determine system performance was done weekly. Each week 10 random wafers were chosen from the wafers which passed through the test facility for that week. Then a copy of the parametric data for all the wafers in the test wafers lot were presented to an expert. From the parametric data, the expert would determine what the test wafers actual diagnosis should be. This was then compared with ADEPT-II's diagnosis. The expert would then rate ADEPT-II on a scale of 1-4 in terms of how well it had diagnosed. On this scale, a 1 represented Exactly Correct, a 2 was Good Enough, a 3 was Not Good Enough and a 4 was Incorrect. Percentage accuracy for each week was measured by taking the total number of 1s and 2s and dividing by the total number of test wafers examined. A chart depicting the accuracy over the first 10 weeks of testing, from Jun. 7, 1991 to Aug. 9 1991, is shown in Figure 4. Over these ten weeks, the week with the worst performance was week 5 with an accuracy of 30%, the best performance was on weeks 7 and 10 at 80% and the average performance over the ten week period was 65.77%. Unusually low performance, such as seen in week 5, was traced back to the fact that the system was not given enough

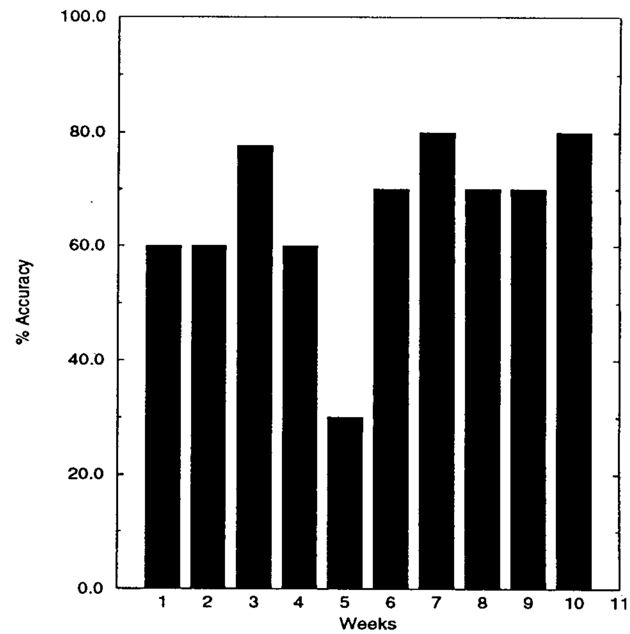

Figure 4: Performance of ADEPT-II Over 10 Weeks.

example support for a specific problem to adequately determine its characteristics. As a result, it incorrectly classified the examples which had that problem, when it encountered them in week 5. Although further testing was performed, beyond the first 10 weeks, it was not done consistently and was not well controlled. It did indicate, though, that on Sept. 6 1991, the system performed at 90% over the 10 random samples for that week.

More informally, it should be noted that the users of the system were quite pleased with its overall accuracy and the system is still presently in use. Presently "GID3_TRAIN" contains 289 training examples. The decision tree consists of 119 branches and a total of 44 distinct problems can be diagnosed. The highest number of diagnoses that the system has had to perform in one day is 200.

Conclusions

Updating the knowledge base of expert systems operating in very dynamic environments can be a very expensive and time consuming task and one which, if left undone, may lead to an obsolete expert system. An alternative to the traditional knowledge acquisition process is to use the approach described in this paper to develop ADEPT-II. ADEPT-II, when installed in this particular domain, was developed more quickly and covered a much larger classification space than the preexisting expert system. The original ADEPT expert system took approximately one and a half man-years to construct and was capable of diagnosing only one product. Modifications which allowed the addition of two new products took another four person-months of effort.

ADEPT-II's creation, initial training and implementation took only four and a half man-months with most of the time savings coming from the knowledge acquisition phase. It quickly reached a level of performance equivalent to the preexisting expert system. After continued weekly training for approximately 3 months, the system reached a performance level of 90% Additionally, ADEPT-II was capable of classifying all product types and was able to make guesses about new products by generalizing from knowledge of previously seen products.

It should be emphasized, however, that while the use of machine learning significantly reduces the amount of time required from experts, the need for an expert still exists. ADEPT-II needs an expert to pick the initial training data for the system and this data must represent a good set of prototypical examples of each class. In addition, the expert is the one who decides which parameters are important for classification purposes in the domain. Later, when new products need to be diagnosed, it is the expert who must first diagnose them correctly so that the system can learn from these new examples.

Acknowledgments

This research is supported in part by grant number 2108111lo from the Florida High Technology and Industry Council and the Semiconductor Sector of Harris Corporation.

References

[1] Cheng, J, Fayyad, U. M., Irani, K. B. and Qian, A. "Improved Decision Trees: A Generalized Version of ID3." Proceedings of the 5th International Conference on Machine Learning. Ann Arbor, Michigan. June 1988.

[2] Joyce, R. J., Perez, R. A. and Sarmiento, C. "Development of a Knowledge Base System for CMOS Parametric Failure Analysis." Proceedings of the Semiconductor Materials International Conference. April 1989.

[3] Koh, S. W. Personal Communication on ADEPT Expert System, Harris Corp. November 16, 1992.

[4] Koh, S. W. Personal Communication on ADEPT Expert System, Harris Corp. December 1, 1992.

[5] Lilkendey, J. T. "Applying Machine Learning to Classification in Manufacturing." Masters Thesis. Dept. Of Computer Science and Engineering, University of South Florida, Tampa, FL. 1993.

[6] Perez, R. A., Sarmiento, C. D., Victor, P. J. and Koh, S. W. "A Modular Expert System Design for Diagnosing Semiconductor Wafers." Proceedings of the Tenth International Conference on Expert Systems and their Applications. Avignon, France. 1990.

[7] Perez, R. A., Hall, L. O., Romaniuk, S. and Lilkendey, J. T. "Evaluation of Machine Learning Tools Using Real Manufacturing Data." International Journal of Expert Systems, Vol. 5, Number 4, pp 299-317.

[8] Victor, P. J. "Machine Learning and Its Ability to Develop Rules For Use Within Domains Containing Classification Problems." Masters Thesis. Dept. Of Computer Science and Engineering, University of South Florida, Tampa, FL. 1990.