Solving Tough Semiconductor Manufacturing Problems Using Data Mining

Robert "Mike" Gardner and Jack Bieker
Motorola Labs
2100 East Elliot Road, MD EL508
Tempe, Arizona 85284
USA
{axsm10, ayvx80}@email.sps.mot.com

Steve Elwell
ON Semiconductor
r13537@email.sps.mot.com

Ray Thalman and Ernesto Rivera
Motorola Semiconductor Products Sector
{rtp730, r15734}@email.sps.mot.com

ABSTRACT

Raising product yield and quality, or quickly solving problems in a complex manufacturing process is becoming increasingly more difficult. Process control, statistical analysis, and design of experiments have established a solid base for a well tuned manufacturing process. However, the dynamic "next-tier" problems such as multi-factor & nonlinear interactions, intermittent problems, dynamically changing processes or installing new processes, multiple products and the sheer volume of data can all make quickly finding and resolving problems an overwhelming task. Data Mining technology applied to data analysis can increase product yield and quality to the next higher level by quickly finding and solving these problems. Case studies of semiconductor wafer manufacturing problems are presented. A combination of self-organizing neural networks and rule induction is used to identify the critical poor yield factors from normally collected wafer manufacturing data. Subsequent controlled experiments and process changes confirmed the solutions. Wafer yield problems were solved 10x faster than standard approaches; yield increases ranged from 3% to 15%; endangered customer product deliveries were saved. This approach is flexible, easy to use, and can be appropriate for a number of complex manufacturing processes.

PROCESSES KEYWORDS

Semiconductor yield enhancement, intelligent data analysis, data mining, machine learning, neural networks, self organizing maps, rule induction, pattern recognition.

1. The Challenge: Debugging Complex Manufacturing Processes

If the world were well characterized, deterministic and linear then our jobs as engineers and scientists would be simple. Unfortunately the world is little understood, chaotic, and non-linear. Real life manufacturing is somewhere in between. Process control, statistical analysis, and design of experiments are critical for the success of complex manufacturing processes such as semiconductor wafer manufacturing, but additional tools are needed. We must constantly face the yield and quality problems stemming from the time lag in our own learning curves as we constantly reengineer our process control for the rapid introduction of new products and new process technology. Today's tools have already allowed us to pluck the "low hanging fruit" of the simple problems. Consequently, business competition forces us to reach for the "high hanging fruit" of manufacturing problems stemming from the complex interaction of multiple factors. Our human comprehension is overwhelmed by the sheer volume of the manufacturing data with its thousands of factors and 10s of thousands of possible interactions. The disciplines of design of experiments is necessary but time consuming. We need additional tools to solve problems faster and solve increasingly more complex problems.

2. The Key: Quickly Formulating the Solution Hypothesis

Semiconductor engineers identify the true causative factors in manufacturing problems using a well known disciplined approach called a "Design of Experiments (D.O.E.)." [5] A designed experiment is an approach to systematically vary the controllable input factors while observing the effect these factors

Table 1: Design of Experiments Cycle Time [7]

Step #	Activity	Typical Cycle-Time (in days)	Cycle-Time Paper Database	Cycle-Time Computer Data
1	State the Objective	1	1	1
2	Select Output Responses	-	-	-
3	Select Input Responses ◄———————————┐	1	60	8
4	Select The Experiment Type and Design	1	1	1
5	Develop Experimental Strategy and Plan	1	1	1
6	Run the Experiment and Collect Data	35	35	35
7	Fit The Model	1	1	1
8	Diagnose the Model	1	1	1
9	Interpret the Model	1	1	1
10	Confirmation Runs Verify Improvement -or-	45	45	45
	Next Experiment beginning again at step 3 ┘	30 iterations	2 iterations	2 iterations
	Total Cycle-Time:	*1276 days*	*202 days*	*142 days*
	Cycle-Time Improvement:	*(baseline)*	*6x*	*9x*

have on the output product parameters. This experimentation methodology is critical in ensuring a statistically significant result. The overall steps in this disciplined approach is shown in Table 1.

The critical step in the entire methodology is step number 3. At this step the engineers must pool their collective knowledge to identify, what they believe, are the most likely causes for the manufacturing problem(s) in question. Each of these proposed causative factors are a hypothesis that will subsequently be tested for validity as input responses in the D.O.E. Unfortunately, this effectively limits the number of hypotheses to only a half-dozen independent causative variables because the number of D.O.E. experiments increases exponentially with the number of variables under test. If one fails to choose wisely, to include the manufacturing problem's true causative factor for testing, all the rest of the D.O.E. work only tells them they missed the mark and must try again. Given that a typical semiconductor process can have thousands of variables and 10s of thousands of variable interactions, choosing correctly can be daunting.

Normally the team selects the potential causative factors by drawing insight from a number of readily available sources:

- what failure analysis or functional tests can quickly be done;
- insight gained from simple statistical examination of some of the manufacturing data;
- engineering intuition developed from years of experience and training.

Frankly, the process is frequently far from objective as individual egos and department reputations are on the line.

In steps 4-6 the engineers design and run the experiments which consists of fabricating a number

of wafers while controlling the causative factors. Of course, fabricating and testing these experimental wafers is costly and time consuming. In steps 7-9 the engineers use statistical analysis to determine which if any of the causative factors were, indeed, the cause. If insightful (or lucky), the experiments bear out that one or two of the chosen factors are actually the problem; corrective action can be taken; and the fabrication of a special set of wafers will confirm the success of the corrective action prior to implementing the changes throughout the line. If, on the other hand, the experiment shows the true causative factors were not found, the entire process must be repeated again at step 3 with a new perspective. It is not unusual for a complex semiconductor fabrication problem to require dozens of iterations which, of course, is a costly cycle-time problem resulting in scrapped product, wasted materials and man-hours on unsuccessful experiments, product shipping delays, lost market windows, and undermined customer relationships.

So, what is needed is a hypothesis generator that greatly increase the odds of identifying the actual causative factors quickly and correctly, the first time. Ideally the lengthy D.O.E. process (steps 4-9) would then be done only once.

3. The Remedy: Using Data Mining

Advances in data mining, manufacturing data collection and desktop computing power are converging allowing for the potential creation of "inductive hypothesis" generators for use in complex manufacturing. A high priority goal for wafer manufacturing is finding the most probable causative factor(s) that discriminate between low yield and

high yielding product by quickly examining the historical data already being collected.

In essence, one uses data mining algorithms capable of efficiently searching very large hypothesis spaces to find the hypothesis that best fits the data. In theory, one could compose a brute force method of testing every possible combination of the collected data variables for correlation against a target variable such as yield. However, the compute time for even a modest manufacturing data set such as in case study #1 would be astronomical. Consequently, inspired "short cut" algorithm approaches that can search the hypothesis space much more quickly for the most probable hypothesis are a must. A number of algorithm approaches exist: [4],[8]

- Decision Trees - inspired by Information Theory;
- Naive Bayes - inspired by Probability Theory;
- Neural Nets - inspired by Biological Models;
- Genetic Algorithms - inspired by Simulated Evolution.

As can be imagined, short cut searches, by definition, may miss certain critical hypothesis buried in the data. Indeed each of these algorithms have built-in biases that potentially "blind" them to certain relations that may be in the data. However, from a pragmatic view point, it's what these methods often can find that makes them worthy of attention. This paper describes in case study form one such data mining method that has been used within Motorola to great effect. [3]

4. Case Study #1: Solving Periodic Yield Dips in a Mature Wafer Line

4.1. Wafer Yield Problem Description

Motorola Semiconductor Product Sector manufactures one particular bipolar logic family in a mature wafer fabrication facility. The process of this fabrication facility was well understood, had a stable process control, and was achieving extremely high average product yields consistently over the previous five years. However, periodically, for short periods of time, the average probe yields would mysteriously degrade by 5% or more. The device failure was usually due to excessive transistor collector/emitter leakage current. Unfortunately, the established process control measurements were not sensitive enough to detect this particular leakage problem until final electrical testing. Consequently considerable added value was lost when the leakage problem was finally detected. In general, periodic yield dips are much more difficult to solve than catastrophic failures. Yield improvements over such narrow

ranges at already high yield are also exceedingly difficult.

4.2. Problem Resolution Through Standard Practice

4.2.1. Previous Studies Performed

Over a period of five years, substantial effort had gone into modeling the leakage current problem. Three primary failuremodels had been proposed to explain the leakage current occurrence. Model #1 proposed that under certain manufacturing conditions a parasitic PNP transistor is formed that allows leakage current to flow directly to the substrate. Model #2 claimed that substantial material stress allows a dopant material path to form creating a leakage path to substrate. Finally, model #3 theorized that a collector to emitter pipe caused by a defect stacking fault occurs somewhere in the process, providing a leakage current path. Device engineers, diffusion engineers, and epitaxial silicon growth engineers performed numerous experiments (>30) throughout this time period to address each of these possibilities with little or no impact on the problem.

Near the end of this five year period, a cross functional team was formed to address the leakage current problem. For over a year, the team performed process comparisons, split lot experiments, screening experiments, and failure analysis in an attempt to determine the cause of leakage. The team compared processing procedure and machine profiles to a nearly identical wafer fabrication line that reported lower yield loss. No significant differences were found. In order to confirm the results of the process comparisons, the team set up and performed process split lot experiments between both wafer lines. Statistical analysis was performed on the process splits to look for significant differences in yield. No statistically significant differences were found. Failure analysis was performed on devices that were failing leakage tests at final probe. Electrical comparative analysis, liquid crystal hot spot detection, photo emission microscopy, and scanning electron microscope cross sectioning/imaging were all performed in hopes of finding a failure mechanism. None was found.

One last Four Factor Designed Experiment, as shown in Table 2, was created and performed in hopes of discovering the key variables influencing the current leakage This experiment happened to catch the transitory process anomaly in operation. The team identified four critical factors that again tied into the original three models for leakage current loss. An experiment was created to determine if any of the factors were significant. Additionally, a test

Table 2: Design of Experiment Factor Table

Factor	Level		Purpose
	Plus	Minus	
1) Alignment	AFA*	Manual Alignment	Tests model #1
2) Pre Clean	Pre Clean	No Pre Clean	Tests model #2
3) Substrate material	Non-Standard	Standard	Tests model #3
4) Post Clean	Clean	No Post Clean	Tests for defectivity

*AFA, Automatic Fine Alignment

for possible defectivity related loss was included. This time the results of the screening experiment demonstrated the use of non-standard substrates to have a substantial yield improvement over standard substrates.

The designed experiment seemed to give credence to the theory model #3 proposed, that a collector to emitter pipe caused by a defect stacking fault occurs somewhere in the process, providing a leakage current path. Due to their relatively small cross sectional area, stacking faults in the 1.4 micron thick epitaxial layer are very difficult to observe from the top after upper wafer layers have been stripped. They are nearly impossible to observe in a cross section. The probable fault mechanism in the bipolar transistors is drawn in Figure 1. A photo sample of stacking faults in a different, much thicker 14 micron thick EPI process is also shown in Figure 2.

This entire investigation took 5 years! The cycle time estimate for 30 iterations of the D.O.E. process shown in Table 1 predicts 4.9 years, which agrees *quite well with the 5 years that* actually had lapsed in this effort.
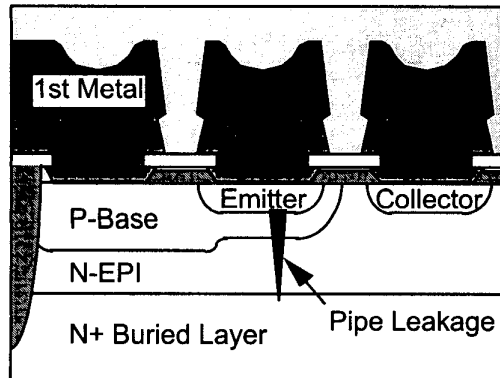
4.2.2. Conclusions and Action

The non-standard substrate material was implemented in the fabrication line on a graded basis beginning in the 3rd quarter of year 1. Additional work on the epitaxial silicon deposition step resulted in the identification and elimination of a problem in 3rd quarter and another in 4th quarter. The implementation of these changes caused a 3% yield improvement in the wafer line as shown in Table 3. The improvement was gratifying, but it took well over five years to find it.

4.3. Data Mining Approach

4.3.1. Blind Analysis Methodology

A decision was made to use an experimental data mining approach. A blind analysis of the wafer yield problem was done in parallel with the device engineering efforts of the device team. Blind analysis forced the data mining results to stand strictly on the merits of the collected data, uninfluenced by the expert's knowledge or the progress being made by the process experiments. To this end, several steps were taken:
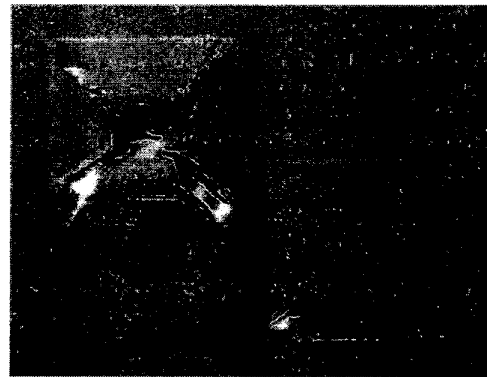

Figure 1: Transistor pipe leakage due to stacking fault.

Figure 2: Photo of stacking faults in 14m thick EPI.

Table 3: Wafer Yield Improvement

	First Year			Second Year
	2nd Quarter	3rd Quarter	4th Quarter	1st Quarter
Wafer Yield	-baseline-	+0.7%	+3.0%	+3.2%

1) The physical significance of the specific data collected from the fab. line was not explained, but simply labeled;
2) The CorDex development team was not given knowledge of the progress of the process experiments;
3) The CorDex development team was composed entirely of process and device "ignorant" members;

Historical wafer data was collected for 2500 wafers from a 2 month period. The input data base measured 133 parameters by 17,246 entries organized into a flat file in Excel. It took 2 months to hand collect, examine and then correct for integrity. The data consisted of all wafer probe data, all process control (PC) data, and selected process step data:

- Wafer Probe Data - The pass/fail count per wafer of 39 wafer probe functional tests were used;
- Process Control Data - There were 59 numerical electrical PC measurements probed at 8 sites per wafer;
- Process Step Data – There were 33 process parameters, hand collected parametric and non-parametric data such as material vendor/lot, wafer boat position, operator & machine IDs, etc.

4.3.2. The Data Mining Algorithm Approach

A combination of self-organizing map (SOM) neural networks [2] and rule induction [6] is used to identify the critical poor yield factors from normally collected wafer manufacturing data. The SOM neural network algorithm performs a type of "multivariate, non-linear regression." The algorithm creates a two-
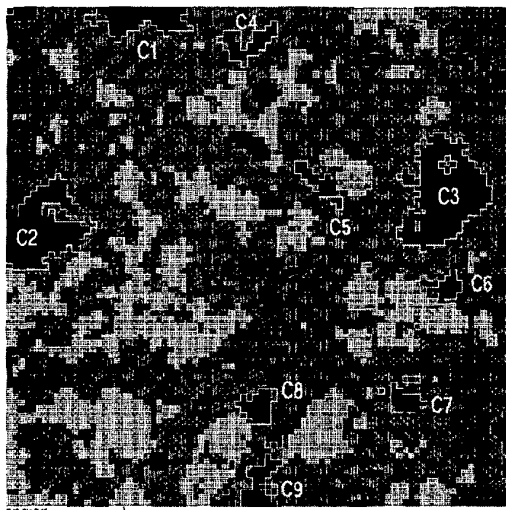

Figure 3: Rule Induction results displayed on cluster map.

dimensional relational topology, called a "cluster map," that best maintains the original 133 dimensional data inter-relationships.

The cluster map for case study #1 is shown in Figure 3 where the black shaded areas indicate low yielding wafers. This is a two dimensional map that shows all 17,246 wafer sites at one time. The two dimensional map is simply a mathematical construct that serves as a trellis by which the data is organized. The X and Y dimensions have NO MEANING whatsoever. The discernment of relevant underlying relationships in the data is displayed by the clusters. The following general rules interpret a cluster map:
1) The presence of clusters indicates meaningful statistical relationships do exist within the data. Random data will form indistinct and/or many small fragmented clusters;
2) The size of the cluster indicates strength of relationship;
3) The number of significant clusters indicates the number of statistically distinct relationships;
4) The spatial relationship of the clusters to one another indicate their relative similarity (spatially close together) or dissimilarity (spatially far apart);

Therefore, significant patterns can be seen as clusters. Special prior knowledge of neural networks or the data itself, is not necessary since the algorithm trains from the historic data without supervision. The algorithm can handle both numeric and non-numeric data and is highly robust in the presence of outlier data. It also works well with incomplete data.

Rule Induction is an additional unsupervised data mining algorithm, that works synergistically with the cluster map. It generates logical expressions (rules) that identify the data attributes that most discriminate between clusters, thus explaining the clusters. Only the classified historic data is used to train the algorithm. After the rules have been generated, they are checked against the original cluster map. The white outlined areas shown in Figure 4 indicate how well the selected rules match up with the original clusters (black shaded areas).

4.3.3. Applying Rule Induction

The case study #1 yield cluster map contains several distinct clusters that indicate significant relationship patterns exist in the original data. The color coding was chosen as follows:
- Black - Poor Yield: the bottom 12.5% low yielding wafers in a left-tailed yield distribution;
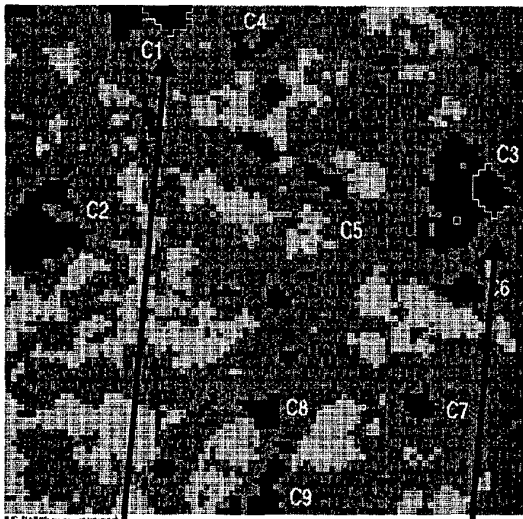- Gray - Medium Yield;
- Light Gray – High Yield.

Figure 4: Cluster Rules C1, #1 and C3, #1 displayed on the yield cluster map.


```
▣═════════ Cluster Rule Analysis ═════════
Clusters:12 Rules:399 (covers 16478 of 17246) orphans:0
► C6, 3 rules covers 46 of 49 items 93% (errors:0)
► C7, 4 rules covers 64 of 65 items 98% (errors:3)
► C8, 4 rules covers 80 of 80 items 100% (errors:0)
► C5, 6 rules covers 62 of 64 items 96% (errors:1)
► C2, 12 rules covers 363 of 364 items 99% (errors:2)
► C4, 5 rules covers 111 of 113 items 98% (errors:4)
► C9, 5 rules covers 137 of 140 items 97% (errors:7)
▼ C1, 7 rules covers 243 of 244 items 99% (errors:7)
▼   1 Rule 680 (9.75) covers:48% (120/1)
      Material Lot = G57494.1W
►   2 Rule 586 (9.53) covers:13% (34/0)
►   3 Rule 498 (9.52) covers:3% (8/0)
►   4 Rule 716 (9.47) covers:22% (56/1)
►   5 Rule 520 (9.06) covers:3% (8/0)
►   6 Rule 584 (9.06) covers:1% (4/0)
►   7 Rule 591 (8.58) covers:6% (20/5)
► Misc.0, 53 rules covers 716 of 739 items 96% (errors:12)
▼ C3, 8 rules covers 291 of 293 items 99% (errors:11)
▼   1 Rule 722 (9.72) covers:19% (56/0)
      Material Lot = G60354.1Y
      YO320 Run = R3521
►   2 Rule 732 (9.72) covers:19% (56/8)
►   3 Rule 734 (9.61) covers:13% (40/0)
►   4 Rule 557 (9.52) covers:10% (32/0)
►   5 Rule 689 (9.52) covers:5% (16/0)
►   6 Rule 695 (9.36) covers:7% (23/0)
►   7 Rule 650 (8.77) covers:21% (72/8)
►   8 Rule 641 (7.19) covers:1% (2/3)
► Hi, 139 rules covers 3751 of 3973 items 94% (errors:83)
☒ consider duplicates when finding matches
[Read Rules]    [Find Match]    [Mike's]    [Close]
```

Figure 5: Cluster Rule Point-and-click Window

The nine largest clusters of poor yield are shown

in black and have been labeled C1 through C9 on Figures 3 & 4. Clusters C1, C2 and C3 are the largest and most distinct. Clusters C4-C9 are smaller and not very distinct from other scattered small clusters. This would tend to indicate that 3 stronger fault mechanisms exist, 6 weaker fault mechanisms might exist, and the remaining are probably due to semiconductor random faults or noisy data.

In previous application experience in marketing and circuit design, the mere inspection of the map provided insight into the underlying relationships within the data. In contrast, the wafer yield problem is much more complex. The cluster process is not sufficient for finding the data relationships by inspection.

This is where the use of rule induction is applied. The labeled data in Figure 4 (original historical dataset, labeled C1, C2,....C9 and other) is processed by rule induction to yield rules or explanations for the clusters shown. The rules are displayed in list form in Figure 5. The significance of this is that now, clusters seen in a map such as Figure 4, can be explained in terms that an engineer could use to pursue a D.O.E. The rules here, visually overlay the cluster map, itself in Figure 4. The rules in black in Figure 5 have been activated for viewing on the map shown in Figure 4, as white outlines.

Cluster C1 is shown in the top left hand side of the map. The outlined portion of C1 is described by rule #680. As an example, the 1st of 7 rules describing Cluster C1 is shown. All seven rules taken together describe 99% of Cluster 1. The rule #680 simply states that 48% of the poor yielding cluster C1 is discriminated solely by the foundational silicon material itself, specifically composed of material lot G5749.1W. Rule #680 is 97.5% confident in this assertion. Similarly, rule #722 of cluster C3 states that 19% of the poor yielding cluster C3 is discriminated by foundational silicon material, G60354.1Y, that has also been through Hipox step YO320, specifically R3521. Cluster 3, Rule #722 is 97.2% confident in this assertion. In this manner the device and process engineers can quickly review the rules as they click through the hierarchy for each cluster.

The accuracy of the rule induction results can easily be seen by activating all rules generated which displays them all on the cluster map. The map in Figure 3 shows this case where all the rules generated are overlaid on the map. Note how well the white outline borders defined by the generated rules match the original borders of the low yield clusters.

4.3.4. Interpretation of the Cluster Rules

Table 4 below lists the primary causative discriminants determined by the induction rules for

Table 4: Case Study #1 Rule Induction Results

Discriminating Attribute	Map #1 (84x84)	Map #2 (60x60)	Map #3 (60x60)
Epitaxial step	38%	47%	43%
Starting Material	13%	13%	23%
Hipox step	38%	0%	0%
Various Other	12%	40%	34%

three separate analysis of the data. The ratio of the individual factors appearing in the rules to the total number of rules is shown in percent. The rules assert that the base silicon substrate material and the epitaxial process steps are the primary discriminants in poor yield lots. The wafers described by the rules containing material and epitaxial references comprises 6% of all wafers and 50% of the poor yield cluster wafers, thus we would expect a 3% yield improvement if both problem areas were fixed.

4.4. Data Mining Conclusions

This data mining approach clearly pointed out EPI and starting material as the critical poor yield factors. The data mining conclusions agreed with the conclusions of the independent process experiments. To a lesser extent Hipox was also identified as a discriminating attribute, but additional collaborating cluster maps were not run. The device engineers used the cluster map viewer to quickly inspect several important parameters of the 17,000+ data entries. In this manner they further confirmed what kinds of physical device structures the fault cluster data represented. From this they quickly grasped the underlying physical fault mechanism at work for major fault clusters. The data mining approach had, indeed, identified the subtle fault mechanisms from the global data given it, and it did it in a timely manner, compared to the 6+ years spent trying to find it through traditional methods.

5. Case Study #2: Solving a Complex Problem in a New Product

5.1. Wafer Yield Problem Description

A second case study concerns a product line at Motorola SPS which manufactures bipolar devices for an automotive application. This line was facing wild variation in transistor beta, putting it both above and below the acceptable limits, which was causing huge numbers of wafers to be rejected then scrapped at probe. This was jeopardizing our customer relationship who was faced with shutting down one of its car assembly lines.

5.2. Problem Resolution Through Standard Practice

Classical theory should have been able to explain this variation in beta, but experimentation lots were run, which confirmed that the classical controls were not the cause of the variation. Additional parametric research and process step commonality studies revealed no reason for the large beta variations. Our data mining method was used to pinpoint the problem, but the result was initially rejected as being impossible. Additional and costly efforts were made to try to solve this using traditional methods, but without success.

5.3. Data Mining Approach

As in Case Study #1, our data mining method was used to cluster the information into statistically relevant patterns; Rule Induction was then used to explain the clusters. Figure 6 shows the map of clusters numbered C1-C4 indicating groups of either high or low transistor beta. In this case, 70% of the data was actually missing from the dataset. Rule induction indicated that a contact layer was the likely cause of the problem. Again, as stated above, the results were initially rejected as a possible area for investigation due to the lack of an obvious known device relationship to the problem.
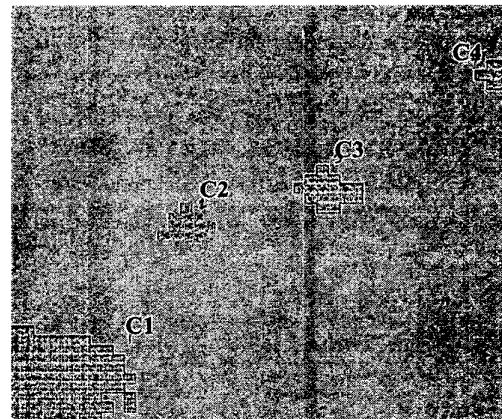

Figure 6: Cluster Map of Beta Groups

5.4. Data Mining Conclusions

Without any success at resolving the problem, this device engineering team proceeded to design a series of experiments to test the data mining hypothesis of contact layer. Much to their surprise and delight, they found a process problem with photo exposure and amount of overetch, causing variations in a contact size. They proved the combination of these two (indicated as contact layer by the data mining approach) were in fact, the root cause of the beta variation. The result of these fixes reduced the average beta by 25% and the variation by 60%. Additionally and most significantly, regaining control of this problem resulted in an increase in yield of more than 15% and ultimately saving the end customer from shutting down their line and finding a new supplier. The data mining approach pointed to the problem in spite of 70% missing data and in spite of how unlikely the true cause.

6. Case Study #3: Improving Yield in a High Yielding Wafer Line

6.1. Wafer Yield Problem Description

Finally, a third case study problem is described as follows. Motorola's Semiconductor Product Sector, Semiconductor Components Group (SCG) manufactures discrete power devices. These are used in automobile ignition applications. Probe yield for this technology, was steady and high for a 16 month period. However, for a three month period, yield performance dropped by 2% with a particular parametric fallout. This is an even smaller anomaly than the case study #1. Other correlating information was that wafer yield maps showed that the concentration of failed die increased toward the center of the wafers.

6.2. Problem Resolution Through Standard Practice

Many cross-functional team meetings involving device engineering, process engineering, design engineering, failure analysis engineering, the contamination free manufacturing team, and management were held to address this issue. Among the courses followed were failure analysis, split lot experiments, starting material evaluation and data extraction. Failure analysis was performed using hot spot detection, SEM with EBIC (Electron Beam Induced Current) as well as Focussed Ion Beam cross-sectioning. Some of these indicated defects, but yielded no real leads to follow to perform a design of experiments. Split lots were run with and without polysilicon back seal (PBS); the ones without

PBS showed a similar circular yield pattern while the PBS ones did not. In an independent and unrelated tool qualification exercise, it was noted that the control tool (T3) exhibited the yield phenomenon while a new tool (T2) did not. It was later discovered that the T2 etcher tool was using a new recipe. Starting material was also evaluated; it was discovered the bulk micro-defect density (BMD) was higher near the edges of the wafers than in the center. Still no significant conclusions were able to be drawn. Finally, process flow comparisons were made between those wafers exhibiting the circular yield patterns and those not, again, though, no explanation emerged for the problem.

While these efforts exposed that there were actual defects and problems, they yielded no real leads for designed experimentation to try to pinpoint the root cause. Further, the only hope of a lead from these analyses, was dismissed and attributed to a change of recipe; since the problem went away, the issue was no longer important.

6.3. Problem Resolution Through Data Mining

6.3.1. Blind Analysis Methodology

When the problem outlined above appeared to have come back again over several ensuing months, it was decided to use the data mining approach to expose any possible leads toward solving this intermittent yield problem. Unlike case #1, the expertise of the production team was drawn upon, as far as determining the type of data to collect. Also, the lowest level of data collected here was only to the wafer level as opposed to sub-wafer for case #1. The following summarizes the data used:

- Final Probe Data - 26 Attributes - pass/fail %, electrical measurements mean and standard deviation;
- Process Step/Flow Data/Equipment - 90 Attributes, equipment, measurements, run numbers;
- Starting Material Data - 3 Attributes, epi reactor, mfg. month, mfg. run number.

Similarly to Case Study #1, data was gathered, by hand, on over 3000 wafers spanning a 2 month production time and organized into a flat file format for a similar analysis method.

6.3.2. Data Mining Implementation

The same data mining approach to analysis was used here as in Case Study #1, though, at this later time, it had been streamlined and improved for usability. As in the previous cases, a data mining expert is not required to use the tools. Neither is it required to have the device/production expert for the analysis phase here, although that person would be able to use and interpret the output of the data mining tools.
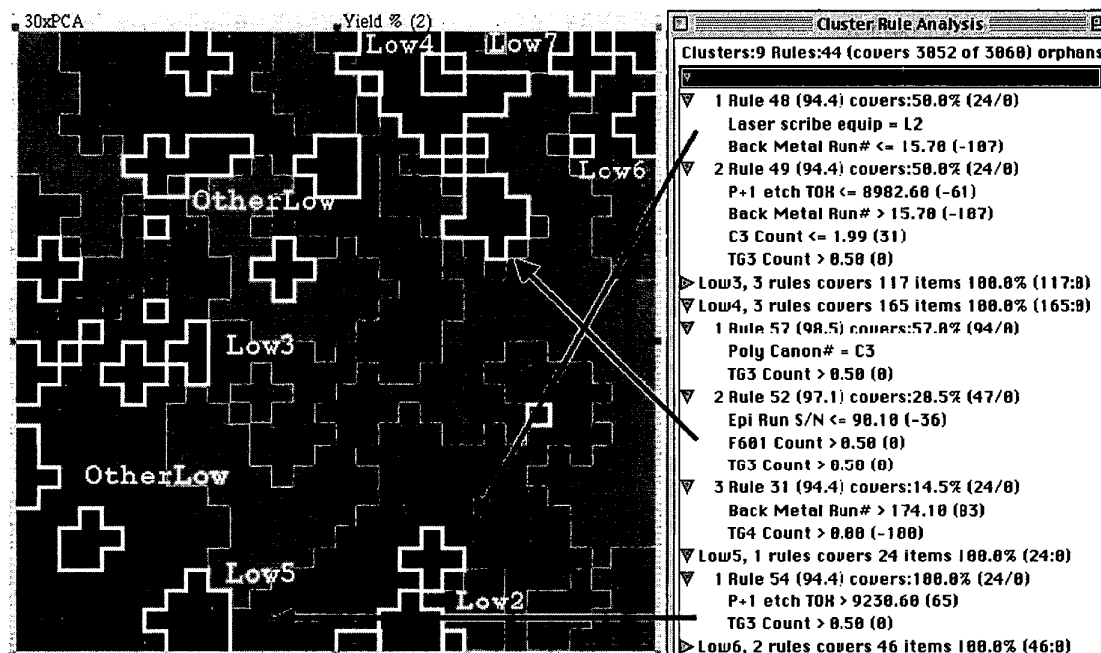
Figure 7: Rules and CorDex Map with Rules Overlaid

Figure 7 shows the result of analysis. Note that the darkest areas represent the visual clustering of the lowest yielding wafers in the database. The listing of the rules to the right in Figure 7 are also shown as the white outlines overlaid on the map to the left. As you can see, there is a very good correlation of the visual organization and the inducted explanation of it.

The rule results indicated that the primary causative discriminants were equipment T3 and T4 for a particular etch step. These were verified in the data using the statistical analysis tools of SAS-JMP; a Tukey-Kramer One Way Anova plot was made for each of the attributes indicated by the CorDex analyses. [1]

One additional feature of the data mining approach (not shown here) is that the same type analysis that is run to find possible explanations for low yield or problem areas can be run to find explanations for desirable conditions such as high yield. This was done for this case and nearly the same set of factors was found to be the most likely causes for high yield; they are simply the inverse conditions. This is an overwhelming indication that the particular piece of equipment is involved in the explanation of the fault mechanism and hence, an opportunity for yield enhancement. Had these factors been different from the low yield indications, additional yield enhancement directions could possibly have been pursued.

6.4. Data Mining Conclusions

The data mining approach clearly showed the T equipment to be a highly likely candidate to pursue in the cause of reducing low yield and ultimately enhancing overall yield. This did agree with an earlier finding where a recipe change/difference was simply accepted as sufficient evidence that the problem had/has been corrected. In spite of this, it appears that until the problem occurs again, these results will be maintained only as a probable avenue to pursue for this manufacturing line.

7. Data Mining Impact on Wafer Fabrication

7.1. Quality Impact

Data mining impacts manufacturing quality on two fronts: product quality and quality of decision making. Product quality is raised because of the correct diagnosis of subtle or formerly "impossibly complex" fault mechanisms. The quality of day-to-day decisions made by device and process engineers can be raised because data mining tools allow them to view large amounts of manufacturing data as a whole. They can more easily see the "forest for the trees" as they have access to a more global view of the process.

Case Study #1: Data mining was able to identify the incoming material type and the EPI process as the problemed process variables. Yield on the line improved 3.2%.

Case Study #2: A 15% improvement in probe yield was achieved.

Case Study #3: The 2% yield dips were eliminated.

7.2. Monetary Impact

Case Study #1: By implementing the three key process changes which resulted in a 3% yield improvement, the cost/benefit analysis studies demonstrated a single year $480,000 savings in non-scrapped die.

Case Study #2: A 15% improvement in probe yield results in savings of $1,050,000 per year .

Case Study #3: Estimated $123,000 in scrap reduction for single events of the intermittent problem..

7.3. Cycle Time Impact

Data mining has the clear potential to drastically reduce manufacturing process improvement cycle time by greatly enhancing the engineers ability to correctly diagnose process problems the first time.

Case Study #1: A cycle time analysis shown in table I shows a problem solving cycle time reduction of 9x assuming a computerized data system is in use. Even a 6x reduction is achieved if data must be hand extracted and assembled.

Case Study #2: It is estimated a greater than 10x problem solving cycle time reduction was achieved. "Actually, it is doubtful we would have solved it at all without the use of (data mining)." - Motorola Product Manufacturing Manager

Case Study #3: Cycle time study was not performed.

7.4. User Impact

Case Study #1: " (data mining) identified the key process variables for the (Case Study #1) problem, independent of the team activity. The tool allows the user to input mixed data types, and reports results in a self organized feature map output that is easy to interpret. This software is highly recommended to anyone who needs to perform screening of large databases for key factors influencing a given response." - Motorola Reliability and Quality Assurance Engineer

Case Study #2: "...we were about to cause the car line to shut down because Motorola couldn't deliver the chips.... So, you saved our ###! (posterior). Thanks a bunch!" "This work is a perfect example of the power of (data mining) – Beta should not have been a function of overetch, but on this product it is – because of a weakness in the process integration. No sensible engineer would have run the DOE he ran, unless he had some hint." – Motorola Manufacturing Manager.

Case Study #3: "I took over the device lines recently. I had little exposure to the issue and the prior non-(data mining) resolution attempts. Yet, working only with (data mining), we were able to identify the probable root cause...amazing! I want to incorporate (data mining) into my standard yield enhancement practices." – Motorola Device Engineer.

8. Conclusion

Finding the causes for lower yielding wafers in a semiconductor process can take many months or years if the problem is complex or intermittent. Since each physical experiment can only test a few variables at a time, the engineer must almost know the precise answer prior to testing, otherwise, dozens of iterative tests must be performed consuming valuable time. Data mining technology applied to data analysis can increase product yield and quality to the next higher level by quickly finding and solving these problems. Wafer yield problems were solved 10x faster than standard approaches; yield increases ranged from 3% to 15%; endangered customer product deliveries were saved. Our approach is flexible, easy to use, and can be appropriate for a number of complex manufacturing processes.

REFERENCES

[1]. Kramer, C.Y. (1956), "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications," Biometrics, 12, pp. 309-310.

[2]. Kohonen, K. (1995), Self-Organizing Maps, Springer-Verlag, Berlin

[3]. Leivian, R., Peterson, W. and Gardner, R.M. (1997), "CorDex: a Knowledge Discovery Tool", in Proceedings WSOM'97: Workshop on Self-organizing Maps, Helsinki University of Technology, pp. 63-68.

[4]. Mitchell, T. M. (1997), Machine Learning, WCB/McGraw-Hill, Boston

[5]. Montgomery, D. (1991), Introduction to Statistical Quality Control, John Wiley & Sons

[6]. Quinlan, J. R. (1993), C4.5: Programs for Machine Learning, Morgan Kaufmanns, San Mateo

[7]. Stuart, Bob (1996), "Motorola Engineering 207 JMP 2 Participant Guide", Motorola University, Feb., p. 1.12

[8]. Weiss, S.M. and Kulikowski, C.A. (1991), Computer Systems That Learn, Morgan Kaufman, San Mateo

[Note: Motorola has licensed it's data mining software to Global Knowledge Systems. For information contact Shawn Smith, ssmith@onr.com]