CHAPTER 3

LITERATURE REVIEW

There are various factors affecting yield of integrated circuit or chip at each process, i.e. from wafer fab process till final test processes. The following is a review of several research studies summarized by the stage of the general process flow at which the study was done.

3.1 Research Done At Wafer Fabrication Process

Francois Bergeret et al. (1999) presented two applications of equipment control using analysis of variance (ANOVA) and nonparametric tests, methods that have been useful in improving yield by detecting problems in etch machines and in reducing defect density by identifying equipment generating too many particles on wafer. Regression analysis is used to quantify the effect on yield of the number of particles measured at different process stages. By means of regression analysis, an objective specification for an acceptable number of particles can be set.

In their studies, ANOVA is used to determine whether there is a statistically significant difference between machines. According to device engineers, more than half of the yield problems is caused by process equipment. The analysis compares the averages of a response variable for the different levels of input variables, and determines if there is a statistically significant difference between these levels. In their investigation the input factors affecting the yield included:

- Three different suppliers of substrates
- Process machines (diffusion furnaces, steppers and etchers)
- The probe test equipment

The response variables for the dashboard's main device (a device very sensitive to yield detractors with a high production volume of more than 30 lots per week) included:

- Thirty-four critical process control (PC) tests, which were used to ensure that the process conformed to the technology, defined by principal component analysis, a statistical technique used to determine the correlation between large numbers of variables.
- Nine electrical testes, which were considered important because they appear first in the Pareto chart of rejects and showed less variation than the probe yield resulting from all the variables in the wafer fab.
- The probe yield, which is the dashboard's main variable.

They found that with a α risk of 5% of the ANOVA, at the implant stage there was a significant difference in yield between the machines. They conducted their analysis by generating 2 kinds of graphs when a significance difference is observed at a given process step: comparative histograms and time slides. Comparative histograms are used to analyze machine distributions. They found that the test of 800 distribution at implant, one of the most important yield detractors, was not normal. Time slides are the graphs used to check causality and plot response variables versus the date of the lots were processed at given step. They are useful engineering plots for determining the date of drift and whether a machine was running alone during a poor yield period. They found that for machine GSD1 which showed a significant reject rate in the comparative histogram, the time slide also showed clearly that this machine had been running only during the drift period.

They also found that a consistent yield difference between two etchers' producing five lots at different times. The results using a DUT board showed a significant 5.5% difference for machine B. As a result, this machine was shut down for several weeks to pinpoint the problem. The output from machine B was analyzed using an exponentially weighted moving average (EWMA) control chart. Preventive maintenance was performed on the machine just before yield recovered – further evidence that the yield loss involved etches. They found that the statistical dashboard could help increase yield by several percentage points.

Given the present state of technology, defectivity problems (such as dust or particles) are the most prevalent cause of yield loss. Thus, they also studied how the particle dashboards can helps considerably to improve yield. Because the number of particles was not distributed normally, they were unable to use ANOVA to test the effect of the equipment on the particles. Hence, they used the non-parametric Kruskal-Wallis test, which compares distribution across K samples without assuming the distribution of the response variable. The particle dashboard alerted them to the fact that one furnace had many more particles than the other. Further investigation showed that the quartz part of the machine was generating particles.

Finally, multiple regression analysis was performed, where the input variables were the particle levels and the response variable was wafer yield. It was used to quantify the effect of particles on yield, define specification limits for the number of wafer particles measured on-line and rank the process steps involved in defectivity. In their finding, all the regressors were statistically significant.

PARTPOLY (particles at poly deposition) was the most significant contributor to yield loss.

Due to the continuing increase in the size of memory arrays, reaching a high yield from the same wafer is more challenging than ever. Nermine (1997) found that redundancy is a way to improve the wafer yield and to reduce the test cost per good die by fixing potentially repairable defects. Redundancy is the process of replacing defective circuitry with spare elements. This is done by programming the right fuses and shifting the array assignments. In Static Random Access Memory (SRAMs), rows and/or columns can be replaced, as well as an entire subarray. Subarray is a unit array of the memory area. The study focus was on yield estimation for block redundancy, a block having a segment of the memory area.

In order to get better estimates of the yield improvement, it is important to realize that not all rejected dice are repairable: a die failing for a short, for example, cannot be repaired. Also the number of defective subarrays that can be repaired depends on the available redundant elements per memory block. This means that having more than one defect per die requires a certain distribution of those defects in order for redundancy to be successful. The yield model used here is Poisson's yield model. Poisson's equation is used to derive a formula for the number of die with a certain number of defects. The formula is used to calculate the amount of improvement expected after redundancy. This improvement is given in terms of the ratio of the overall good dice per wafer to the original good dice per wafer after considering some key factors. These factors are memory area, available redundant elements, defect density and defect types with respect to the total reject die and defect distribution on the memory area.

Chou et al. (1998) set up a control chart for the number of particle counts on wafers. Particles on the wafer surface can lead to failure of the electronic circuit and loss of yield (or percentage of acceptable items). Particles deposited on a wafer surface can be generated by several sources; the manufacturing environment, including people working in the manufacturing area; by the equipment used to manufacture the wafer; and by wafer handling. In particular, equipment-generated particles can be one of the main causes of yield loss. Particle control is a major factor for yield, quality, and reliability. Therefore, quality improvement of the manufacturing process is of the utmost importance.

The purpose of the analysis is to find out whether the process involving the selected equipment is in control and what areas need improvement. Particles generated by the process equipment generally are caused by mechanical operations such as wafer handling, wafer rotation and axis tilt, pumping and venting, and valve operations. To minimize particles from these sources, the equipment must be routinely monitored to ensure proper operation. A non-production surrogate or test wafer is an unpatterned wafer that has been analyzed by a laser-scanning device and found to contain few or no particles. In their study, they found that transforming the original variable (number of particles per wafer) to a normal variable yielded the best fit. In the study, they have considered several transformations including Johnson system and the logarithmic and square root transformation.

Alvarez et al. (1988) used a Box-Behnken experimental design to optimize a BICMOS process. After obtaining a set of second-order equations relating the

various process steps and the device electrical characteristics, they optimized the process by determining the range of process parameters over which all of the responses meet their specifications. However, they used 1-D device simulator for their analysis, which is not appropriate for our analysis, as the inclusion of 2-D effects is mandatory for determining device characteristics of submicron devices.

Low et al. (1989) used experimental design to build a macromodel to analyze a 3 μ m CMOS process. In their paper, the authors used a reduced number of experiments in order to screen the significant input factors before performing the simulations for the full experimental design. Then simulations were performed considering only the significant input factors. The approach described by the authors is extremely useful for problems with a large number of input factors (> 10).

3.2 Research Done At Assembly Process

Gooi (1993) analyzed the possible causes of high machine-to-machine variation in the wire bond process and then presented an approach to bring about a reduction in this variation at wire bond process. He identified a few factors of machine-to-machine variation under four categories: -

a) ***Man*** - *variation in set-up and maintenance by different operators*

b) ***Materials*** - *incoming material variation and differences made*

c) ***Methods*** - *capability of measurement methods*

d) ***Machines*** - *differences in machine parts and input parameters set at different levels*

He found that the causes of machine-to-machine such as variation in incoming materials, variation in machine parts, variation in set-up and maintenance by operators can be readily resolved by the development and effective application of proper procedure, better control, better training and better awareness program. This leaves two other key items that require special attention: development of capable measurement methods for key input parameters, and the standardization of key input parameters for all machines. The wire process have several key input parameters and among them are usually programmable bond power and bond time, initial bond force and temperature.

He had identified five key input parameters, which are critical in maximizing the energy transfer from the ultrasonic generator to the wire during the bonding process. These five parameters are wedge/capillary tightening torque, transducer clamping torque, initial bond force, heatsink/post clamping force and generator calibration. Since these five input parameters are critical, they have designed and developed better measurement methods for these parameters. A torque wrench and torque watch is introduced to enable quantitative measurements of the wedge/capillary tightening torque and the traducer clamping torque respectively. Regarding the measurement of initial bond force, a vertical linear motion jig was designed and used to eliminate the read to hold the gram gauge and consequently reduce the variation due to hands-on handling. The vertical linear motion jig not only enables capable measurement of the initial bond force but was also used in the measurement of heatsink/post clamping force.

After the development of capable measurement methods for the identified key input parameters, he then collected data to identify the best performing machine. Data on two critical responses, the wire pull strength and the ppm defective level, was collected from all the machines and then pareto to determine the best performing machines. He then used the developed measurement methods and captures the levels of the key input parameters of the best machine. He standardizes of the levels of these five key input parameters across all machines by benchmarking best performing machine and fan-out to all the other machines. The assumption here is that if the best performing machines can performed well under those levels of key input parameters, then it is possible for the other machines to do equally well with those levels of key input parameters.

The result achieved in this application is significant; not only the machine-to-machine variation is reduced but also the overall yield loss in wire bond process is improved from 1500ppm to about 750ppm-yield loss.

3.3 Research Done At Test Process

Teh and Wong (1998) discuss a new technique in the final testing of semiconductor devices, i.e. Statistical Method Achieving Reduced Test-time (SMART), an auto-skip testing method to optimize final testing. This SMART test was applied to matured and stable devices, which do not have any wafer process changes. As such, there is not much opportunity for final test yield improvement, except to sustain the yield at its current level. Thus to reduce the final test cost (yield loss) further, the SMART test was carried out to improve the test system

throughput, using statistical methodology to reduce test cost through test time reduction.

Based on the historical data distribution for each parametric test, they identified the parametric tests that have very good process capability (with six-sigma limits). Parametric tests are performed on some of the specially designed test structures on each die (in the scribe area). These test measure factors like threshold voltage, resistance, capacitance, etc. These identified parametric tests will be candidates for the SMART test.

Their idea is to test a sample at the start of every lot and perform all open/shorts, functional and parametric tests. Upon completing testing with this initial sample, they can predict the process capability of each parametric test identified as a candidate for the SMART Test. This process capability will be compared against a set of six-sigma based criteria. If a parametric test meets the criteria, it then qualifies to be skipped for testing to the remainder of the lot. Otherwise that parametric test will be tested 100% as normal.

After the above comparison has been done for all the identified candidates, testing will continue with reduced number of tests depending on how many candidates meet the criteria. As a result, test time for the lot will be reduced. They concluded that SMART Test (test time reduction) could reduce final test yield without compromising on product quality.

Chan and Ng (1999) found that the ZBTRAM[1] was having high marginal rejects failing at I/O[2] level test, T81. The marginality problem found on the device testing forced the rejected units to be retest. Many rejects were found to have failed T81 (high I/O levels test) for ZBT devices (MCM63Z736TQ, MCM63Z737TQ, MCM63Z818TQ, AND MCM63Z819TQ). Until WW46 1998, there were about 5% yield lost due to T81 failures. The debugging process of T81 I/O level failure was complicated. It involved split input and output channel testing methodology to examine its very fast (150MHz) no dead cycle switching in between READ and WRITE operation. They found that these unique characteristics of ZBTRAM were the main reason of having high T81 rejects compared to other device based on the same technology, like the 0.35UM Synchronous Burst RAM (MCM63F733TQ).

In their study, they found that for x18 devices, a functional unit will consistently fail for site 9,11,13 and 15 but passed at 10,12,14 and 16 for tester handler #2. DTLOG[3] result showed that all sites failed with same type of mechanism. Functional DTLOG logged from site 9, 11, 13 and 15, showed that they were failing at the same I/O pin. The possible reason for this type of failure that the driver of the PE card in the tester, which was used to drive the signal for all 4 respective sites (9, 11, 13, and 15) was having problems. In this particular

[1] ZBT gained the name from zero bus turn around, where, there is zero dead cycles or "no-bus-latency" during its transition from READ, WRITE to READ. This type of design offers 100 percent bus efficiency for user and this feature is particularly useful in a high bandwidth demand application like router or bridge in networking.

[2] I/O: Input Output. Device or terminal of networks which combines input and output with one terminal.

[3] An Advantest Test System Utility that helps user to retrieve the current/voltage value (parametric) measured a write/read operation (functional) performed in a particular test subroutine.

PERPUSTAKAAN UNIVERSITI MALAYA

case, they found that "P45" (DQB06, an I/O pin) was the problem. It was confirmed that this failure is not temperature dependent.

In analysis of verification on hardware, they swapped DUT boards, test head, pin cards, and tester to confirm that even with hardware to hardware variation, the T81 problem remained. The problem of T81 could not be solved after some analysis on hardware. A similar analysis was then carried out with the Intel tester (T5581 model). The same failure resulted, and further testing with different testers confirmed that the problem was not due to the tester.

Finally, they found that the device was very much sensitive to the setup and hold time, which were 5 nanosecond and 1 nanosecond respectively during analysis on software and timing parameters in the test program. With longer setup and hold time adjusted in the new test program, units were more reliable at I/O test level necessary QC failure had also been eliminated. This solution had improve yield by 30%, i.e. from about 62% in WW40 1998 increase to 92% in WW12 1999.

3.4 Summary

This chapter group studies into three areas i.e. research done at wafer fab process, assembly process and test process. These researchers have investigated the effect of machines, man and test procedures on yield. The methods used are a variety of statistical techniques, some applied to historical data, other to experimental data.

The analysis approaches described by these authors include:

a) Analysis of variance and non-parametric methods in order to detect problem equipment and reduce defect by identifying problem equipment.

b) Poisson's yield model to derive a formula to calculate the amount of yield improvement.

c) Data transformations such as logarithmic and square root transformation.

d) Pareto chart to determine the best performing machines based on two critical variables collected, i.e. wire pull strength and the ppm (part per million) defective level.

This study evaluates yield obtained at the test process using descriptive techniques, ANOVA, non-parametric methods, multiple comparisons, histogram and Pareto charts.