FACTS AND FICTION ABOUT THE RELIABILITY OF ELECTRONICS

Örjan Hallberg

Ericsson Telecom AB

S-126 25 STOCKHOLM, SWEDEN

Abstract

This paper reviews the reliability of electronics within the telecom area. Assumptions used for prediction of the reliability of an electronic system are reviewed and compared with information gathered from the largest computer system in the world - the telecom network. Methods used for analysing field reliability data are described and a new method using parametric optimisation is presented.

THE RELIABILITY OF ELECTRONICS is becoming increasingly important Hospitals, banks, traffic control telecom business, etc. all depend on reliable electronics. Today the quality requirements on individual components or building blocks are so high that it is difficult or even impossible to verify failure rates during the early qualification or later in the production phase. Hence, there is a need to analyse the quality and reliability of the products based on real field use data.

It is also necessary to review the assumptions made for the system reliability prediction. As the quality levels improve, we have to improve our methods of quality prediction.

In this paper we will look at some of these assumptions and compare them with findings from real field use data. We will also look at different methods for analysing field failure data.

THE LIFE CYCLE COST

Today we are used to having more or less failure free electronics. The banking machines always work, the telephone lines and power lines work most of the time. Car electronics is getting more complex and at the same time even more reliable.

Several conferences are addressing different reliability topics. Just to mention some of them:

Reliability Physics
ESREF
ESD
ISTFA
Test structure workshop
Integrated reliability...

Many books and articles can be found on the subject.

The reliability of a device or a system is also an important part of the total Life Cycle Cost (LCC) of that part or system. Failure costs are multiplied by a factor of at least 5 for each step of the product life cycle that is passed. Important parts of the LCC are e.g. failure costs at the board testing, failure costs at the integrated system testing, failure costs at the installation testing and failure costs incurred in the field use during the guarantee period or later on. We can see a correlation between these costs in our systems. Well designed and assembled products often cause low costs all the way. See figure 1 that gives a summary of statistics from a set of different mass produced boards. The numbers have been normalized but the relations between them are correct for this study.

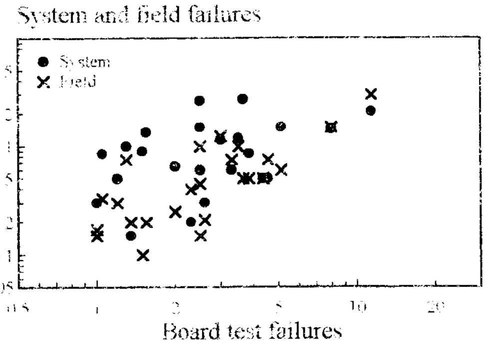

Figure 1
The correlation between board test failure levels and failure levels at system test and in field service.

One major part of the LCC is the test cost. It is clear that improved reliability and reduced failure levels will open up for reductions of the test costs. This can e.g. be done by sampled testing, Ship to Line (components), shorter test times, test programs that adopt to the statistics by themselves etc.

This shows how important a good product reliability is even without considering the negative commercial effects that a bad reliability reputation can and will cause.

In this paper we will mainly be looking at microcircuit and board reliability. The methods presented can be used to evaluate the reliability of any product.

HW RELIABILITY REQUIREMENTS FOR A TELEPHONE SWITCHING SYSTEM

The subscriber expects to complete calls without interruption. In reality there is a small risk that he will experience a certain downtime per year. The reliability of each single electronic switching system is the key element of the total network's reliability. It is designed to have an availability that often is described by these characteristics:

- System restarts (number/year)
- System down times (measured in minutes/year)
- Line down times (measured in minutes/year)
- The number of subscribers disabled due to a single HW failure must be limited
- HW redundancy is provided for all equipment handling more than 128 lines.
- HW failure intensity is often measured as the number of board replacements per 1000 subscribers and year.
- Individual boards and components are measured by MTTF and FITs where MTTF stands for Mean Time To Failure and FIT is a short for Failure unIT, equal to one failure per 10^9 component hours.

FIELD RELIABILITY MEASUREMENTS

Within the telecom area there are several methods applied for the analysis of field reliability data. Some of them will be described in the following.

AVERAGE FAILURE INTENSITY

This method assumes that the failure rate is constant so that the failure rate and the failure intensity is the same. This is the same as assuming that the HPP model is valid (homogeneous poisson process=HPP)[1]

In this way the amount of boards or components put into service are multiplied by the field use time in order to get the total number of component hours. Then the total amount of registered faults is divided by that number to get an estimate of the failure intensity.

This method is still frequently used because it is simple. It has, however, a few drawbacks. First, it accumulates numbers, and, will in the long run be fairly insensitive to detect increasing failure rates. Second, it gives no information on early failure rates of the products.

NON PARAMETRIC ANALYSIS, NPA

The method of non-parametric analysis is described by J. Moltoft[1]. This method has successfully been used for a number of years at Ericsson Telecom AB. The main reason behind this was that we needed to quantify both the initial failure intensities and to be able to detect any sign of early wear-out preferably well in advance of our customer. As our hardware population (e.g. AXE switching systems) is continuously growing we need to be able to analyse the data taking the growing population into consideration.

Non parametric analysis does not depend on any presumed life distribution of the studied object. On the other hand, it requires a great deal of data before it can be applied. The following is required:

- The time (date) when each unit is brought into service
- Identification of the location of each unit
- The number of units installed at each location

A repair report for every individual returned part stating at least the following:

- Product identification number
- Location where it failed
- Failure date
- Failure description
- Repair information

This information has to be collected for each market that we want to cover, either totally or from a representative sample of AXE systems.

Figure 2 is an example of a study of one board type. The graph shows the number of installed boards and the number of reported faults per each month.

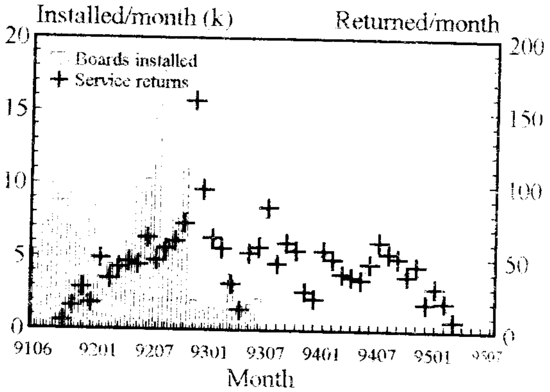

Figure 2
An example of field failures from one board type installed during two years and followed for four years.

All 150k boards have today been working for at least two years and a few up to five years.

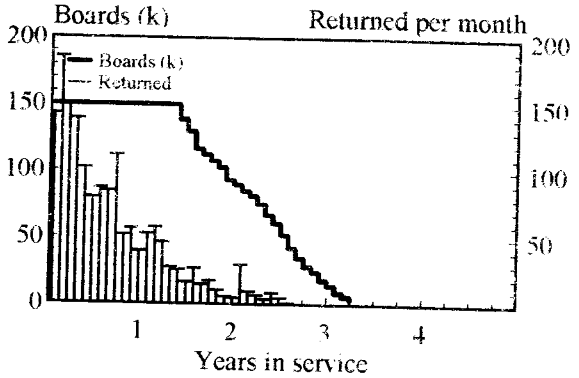

Figure 3
The number of failures per month of use time when all installed boards are merged into a common start time.

The information from the repair reports was analysed at the end of 1994 by using the non parametric method. Each report was traced back to its location and service start time. A computer is then used to compile a population file where all units used for the analysis are assigned a common start-time. The difference between the failure date and the start date then gives the field use time for each failed device. The number of devices in active service is then plotted as a function of field use time, not calendar time as in figure 1. Figure 3 that gives both the number of devices in service and the number of returned devices vs use time. When this analysis was made all parts had been in use for at least 1.5 years and some of them up to three years.

This information is then used to compute the failure intensity of the product of interest in each censored time interval e.g. month. Figure 4 shows the failure intensity and accumulated failure level of the product that had the service profile of figure 2.

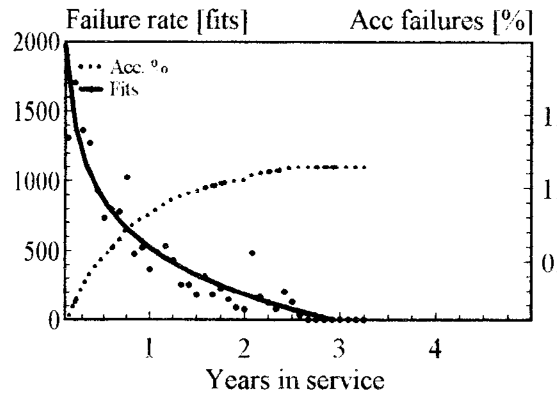

Figure 4
The point estimate of the failure intensity per month of field use time and the accumulated failure level in % of the base population that has been in use for that time.

This method is used both for board and component reliability analysis. In this case the failure intensity is calculated as a point estimate where the number of faults are divided by the number of units that have been in use during the censored time period. More examples are given later on.

The main draw-back of this method is that it requires a lot of detailed data that often can be difficult to get hold of.

PARAMETRIC OPTIMISATION ANALYSIS, POA

Most of the manufacturing industries know how many parts they have sold each month or each year. It is also common that they have information on the number of products that have been returned back for service or repair, at least during the guarantee period.

It is less likely that the manufacturer can trace all failures back to the individual user and to the point in time when each particular product was brought into service. In order to be able to make use of all available repair information we have developed a general procedure for field data analysis with minimal requirements on the raw data.

Table 1 gives an example of such a set of minimum field data information. It states only the censored time intervals (months), the number of units brought into service by each month and finally the total number of returned units by each month. The table does not say from what installation period one particular returned unit comes from.

The aim for the analysis is to find a life distribution F(t) for the studied product that gives the best fit to the measured data such as the fictive example given in Table 1.

Year/month	Installed	Returned
9312	1000	9
9401	500	9
9402	2000	25
9403	2250	35
9404	500	26
9405		20
9406		16
9407		14
9408		13
9409		12
9410		11
9411		10
9412		10
9501		9

Table 1.
An example (fictive) of minimum information needed for the analysis of field reliability data using the POA method.

T	N	T_1	T_2	T_3	T_4	T_5
T_1	N_1	N_1F_1	N_1F_2	N_1F_3	N_1F_4	N_1F_5
T_2	N_2		N_2F_1	N_2F_2	N_2F_3	N_2F_4
T_3	N_3			N_3F_1	N_3F_2	N_3F_3
T_4	N_4				N_4F_1	N_4F_2
T_5	N_5					N_5F_1
T_6	N_6					
T_7	N_7					
.	.					
.	.	r_1	r_2	r_3	r_4	r_5

Table 2.
The calculation of expected failures during each time period T_i. For each period N_i devices are brought into service. F_i is the probability that a device will fail in the interval T_i of its own service time. The sum of each column is compared with the actual number r_i returned during each period.

Definitions:

T_i = Calendar time period i
t = Field use time
N_i = Number of units put into service
r_i = Number of returned units during period T_i
$F(t)$=The cumulated failure density function
$F_i = F(t_i)-F(t_{i-1})$= the probability of failure in period T_i

Table 2 shows in a matrix how the expected number of failures are summed up for each time period. The game is then to find the function F(t) that gives the best fit to the real amount of returned devices.

This calculation is performed by an Excel 5.01 application where an automatic optimisation subroutine is used. Basically the sum of all squared differences between measured and calculated (accumulated) failures are minimised. The application has at present the ability to use four different distributions: Exponential, Normal, Lognormal and Weibull.

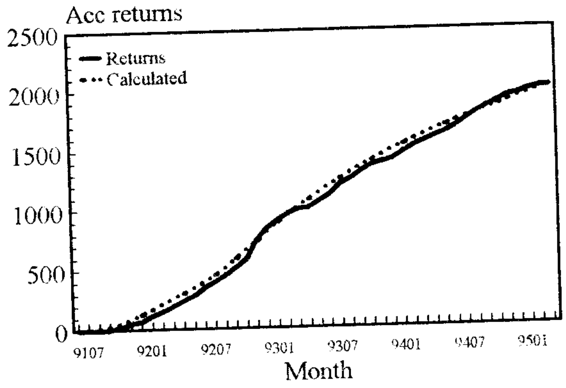

Figure 5
Result of a parametric optimisation analysis of the data given in Figure 2. The accumulated number of faults are calculated by the optimum F(t) function.

An example is given in figure 5 where the accumulated number of returned devices is calculated and compared with the reported numbers. The optimal distribution is then used to plot the cumulative density function and the failure intensity. This is done in figure 6. The POA calculated result is based on a larger amount of raw data for a longer time than the non parametric analysis was based on. This explains the small difference seen in the results of figure 4 and figure 6. We believe that the NP results indicated a too optimistic projection due to lack of data at the time the analysis was made.

FACTS AND FICTION

We now have the possibility to investigate the large amounts of reliability data that continuously is being collected from many thousands of AXE telephone switching systems that are operating around the world. We have detailed data representing more than 38 millions microcircuit years available in our databases. Next, we will take a look at some of the general assumptions that often are made when the reliability of electronic parts are predicted or analysed. Results from our field reliability studies will show how relevant these assumptions are.

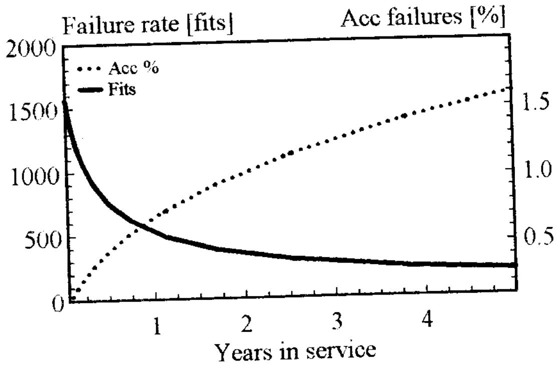

Figure 6
The failure intensity and accumulated failure level vs use time as calculated from the POA method

Reliability prediction

Is it really useful to predict the failure rate of complex systems just by adding the predicted failure rates of its components?

Our experience with part count based reliability prediction is not that good. Figure 7 gives a scatter plot of the predicted failure rate and the measured average failure intensity of 50 different board types. The prediction ranged between 100-5000 fits while the measured values mainly fell within the range 3-1000 fits. No clear correlation is seen.

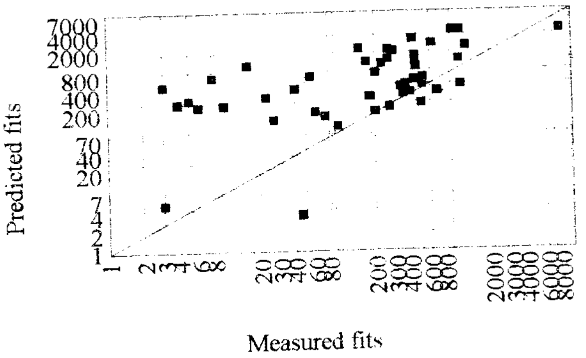

Figure 7
Scatter plot of predicted and experienced failure intensities of 50 different board types. The plot does not indicate a good correlation.

Board reliability depends on a lot more than just the component reliability. Main repair codes are:

- Component replaced
- No failure found
- Repair not finished
- Other reasons

A better way seems to be to follow the failure statistics per each board type and to calculate the needs for spare parts based on that statistics rather than on the sum of component failure rates. A new way of reliability prediction might be developed by looking for a correlation between the measured reliability and different board characteristics apart from its component population.

Is the failure rate constant in time?

In order to answer this question we used the NP method to analyse the failure intensity over time for a few component classes. Figure 8 shows the typical behaviour of four classes of microcircuits. Analogue/interface/bus circuits have a clear time dependency. Bipolar logic has such a low failure rate that it can be difficult to make a plot. DRAMs of 64k and larger seem to show a constant failure rate.

Knowing that the main destructive failure mechanism for DRAMs is an oxide breakdown the constant failure rate of DRAMs should not be a surprise and was as a matter of fact postulated in a report of 1MDRAM qualification tests already in 1990[2]. The high voltage screening that is performed by the manufacturer is equivalent to a very long time at normal use conditions. An 8h stress at 3.5V/100Å for 1MDRAM was estimated to be equivalent to 3.7 years at normal use conditions. The effect of such a 'long' screening is that the failure rate will stay constant for roughly the same amount of time in use, i.e. for a number of years.

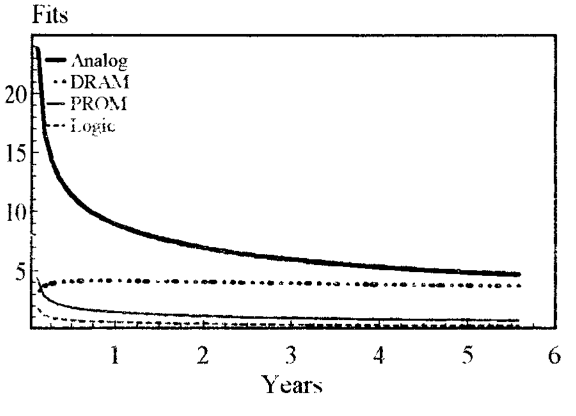

Figure 8
Failure rates of some IC families plotted as a function of field use time. The data was analysed by the NPA method

When looking into the failure intensity of boards we can see a clear time dependency in most cases like in figure 6. In average the return rate of boards will be reduced by 50% every 3 years. ('half life')

In summary, the failure rate is not constant but steadily decreasing over time for most well designed electronics with the exception of well processed and deeply screened memory products that show a constant failure-rate.

There is also an other time dependency of the reliability and that is due to the learning curve of components . As the volume produced accumulates so does the production knowledge. Figure 9 gives a summary[3] of the development of microcircuit reliability over the past 20 years. The DRAM data and data plotted at 1990 is from our own experience

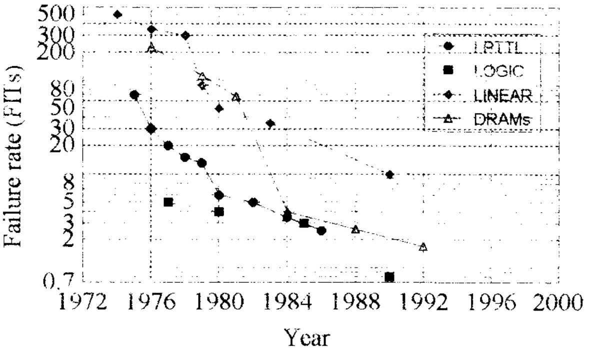

Figure 9
The improvement of the IC reliability over the years due to technology development and quality improvement.

What does burn-in do to the reliability?

A short burn-in or voltage stress screening simulates a longer field use time at normal conditions. A deep screen will then result in a failure rate that is fairly constant for quite some time. During that time the failures will accumulate linearly as shown in figure 10.

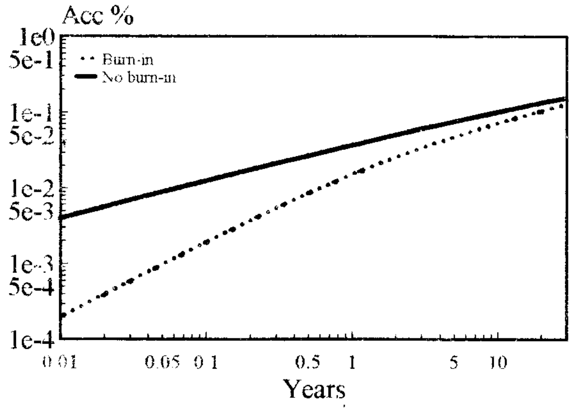

Figure 10
Accumulated failures over time from a non screened population and one that has been screened (bottom line)

In this example one can see that the curve bends asymthotically towards the stright non screened line. In this hypothetic case the screening was similar to almost one year of normal use. If one comes across a similar result where the slope of the failure accumulation at the beginning is linear by time it is most likely that the product has been stress screened.

Thus, a bent curve like the one in figure 10 doees not necessarily indicate a weak subpopulation but just that the main population has been screened.

Is there a complexity dependency?

Traditionally, when predicting the failure rate of microcircuits, it is assumed that the failure rate depends strongly on the complexity of the device. As the board reliability prediction so far has only been based on the part count method a similar procedure has not been developed for board level application. This may, however, be an area for future study.

Every new VLSI process generation seems to require a new manufacturing facility. In order to get a profitable product the process control and defect density must be steadily improved. This process improvement has so far more than compensated for the increased chip sizes and complexity. An example is shown in figure 11. The measured failure intensity of DRAMs is decreasing despite the fact that the complexity has increased by a factor of 1000 in terms of bit size.[4,5]

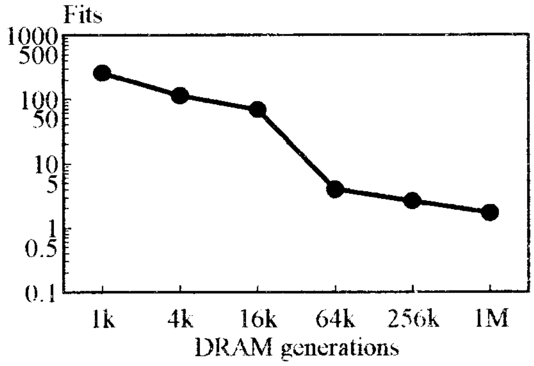

Figure 11
The reliability of DRAMs of different complexities as an average failure rate over two years of field use.

The only complexity dependence that we have seen so far is that the failure intensity will decrease as the complexity increases. This is contradictory to what has been postulated based on the assumption that the defect density is the same on ICs of different complexities and chip sizes.

Is a hermetic device more reliable than a plastic packaged one?

Up until about 1986 plastic packaged ICs were not really accepted in our systems. Tests performed around 1970 had given bad results and it was not easy to change the general attitude against the use of plastic packaged ICs in telecom applications

In 1983 we started to take an other look at the reliability of plastic packaged ICs and we could see indications of improved performance, especially in terms of moisture resistance.

The most important package type for ICs was earlier the CERDIP or CDIP that basically is made up of two ceramic plates sealed by a glass material that melts around 450 °C. The main advantage was that the package could be seen as hermetic and that the chip would be perfectly safe inside that package. Unfortunately there was also a risk of trapping a lot of moisture inside that package during the sealing process. Around 1975 there were several reports on problems with corroded CMOS circuits found inside hermetic CDIPs.

An other problem with the CDIP was that it could leak or sometimes first start to leak and then be plugged hermetic again. A non hermetic CDIP can cause either corrosion or e.g. functional problems due to electrical effects.

A package is traditionally considered as hermetic if it meets the hermeticity standard of $5*10^{-8}$ cm^3 Atm/s. On the other hand it can be shown that such a leak rate will fill the package from the surrounding atmosphere within a few weeks. What we need is a package that is capable of holding the cavity ambient intact for 20 years or more. For that purpose we would need a hermeticity limit of something like $2*10^{-11}$ Atm/s.

In order to evaluate if CDIPs really were that airtight we developed the 'ultimate hermeticity test' (UHT). As CDIPs are sealed at 450 °C the cavity pressure will be only 0.5 Atm once the temperature is back to normal. The amount of gas inside is then equivalent to 12 mm^3 at 1 Atm if the package is hermetic. If it leaks the amount of gas will be around 25 mm^3. The UHT method was simply to break the package into two parts under the surface of silicon oil and to collect the gas bubbles into a capillary tube so that the gas amount could be measured. Figure 12 shows statistics from CDIPs that have been rejected and accepted by a standard hermeticity test.

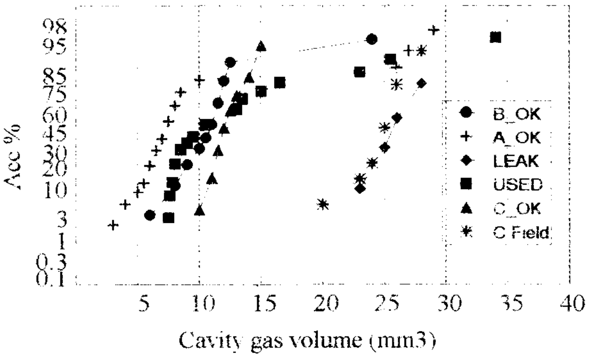

Figure 12
The distribution of cavity gas volume inside CDIPs measured by the UHT method. Airtight devices have around 12 mm^3 of gas while the leakers always contain around 25 mm^3.

As much as a few % could be found to have leaked at least once in their life. Parts taken from sockets after six months use on printed circuit boards did show a wider spread and at least up to 40% were affected in one specific case. Investigations showed that socket insertions and temperature cycling typically could introduce 3% of leaky devices at the time when the tests were performed. (1979, 1982).

In a few cases we have seen CDIP packaged devices failing in the field due to this kind of unhermeticity. The failures were caused by MOS parasitic action due to charge spreading on top of the chips. All failing devices that were tested by the UHT method showed clearly that they had leaked. See figure 12.

Since plastic encapsulated ICs become accepted, we have now the possibility to compare the reliability of CDIP and plastic packaged devices of those types that have been produced in parallel for some years. Figure 13 shows the field reliability of one such device type that was produced in volumes both in plastic and CDIP versions. Both versions have a decreasing failure rate but the average FIT rate is in favour for the plastic packaged one by 2.2 to 12.9 FITs.

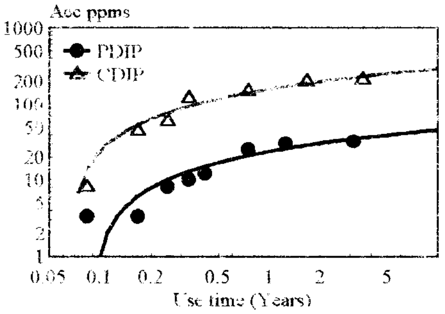

Figure 13
Field reliability data in terms of accumulated failures in ppms for one microcircuit packaged in both CDIP and plastic encapsulation. The average failure rates are 12.9 and 2.2 fits respectively.

As to our knowledge, the reliability of plastic packaged microcircuits is as good as or even better then the reliability of 'hermetic' counterparts today.

Is there a temperature dependence of the failure intensity?

Traditionally, failure rates are predicted based on the assumption that there is a temperature dependence. In many reliability reports from users and manufacturers such a dependence is used in order to correlate the findings from accelerated tests to the normal ambient. The temperature dependence is often characterised by an Arrhenius function and the key factor is the 'Activation Energy', measured in eV.

The temperature dependence used in prediction models like MIL HBBK 217 or in our internal data base is fairly low, around 0.25 eV for microcircuits. But when accelerated tests are going to be evaluated the numbers used are more like 0.7 or 0.9 eV that gives a much higher acceleration.

One basic problem is that it is not at all relevant to talk about an activation energy of failure rates unless it has been shown that the failure rate really is caused by an exponential life distribution where the failure rate is constant over time. It is only relevant to talk about the temperature dependence of the life distribution, e.g. the time to 0.1% or to 50% failures.

If the failures are mainly caused by a sub population of, say 1%, then it is relevant to characterize that sub distribution and the temperature dependence of its median life, i.e. the time to 0.5% failures.

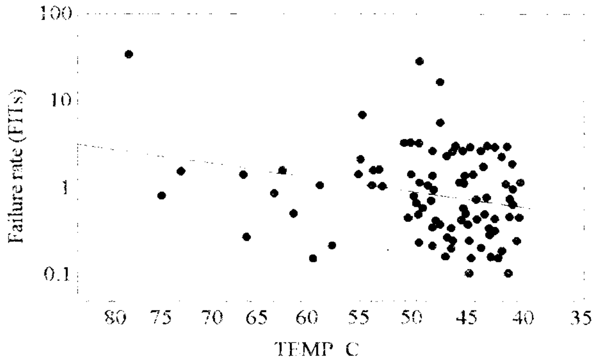

Figure 14
An Ahrrenius scatter plot over the individual average failure rates measured on different bipolar logic ICs. The median failure rate is around 1 fit and no strong temperature dependence is noticed.

Figure 8 showed the failure rates of four different microcircuit families. The average temperature of the analogue family is higher than the logic family but this is not the main reason behind the difference. We looked into the reported return rate of a number of circuit types within the bipolar logic family in order to see if there is a trend or relation between reported failure rate and estimated chip temperature. Figure 14 is an Arrhenius diagram that plots the individual failure rates for each IC type versus estimated chip temperature. The computer seems to believe that there is a slight temperature dependence that can be characterised by 0.33 eV. The only conclusion the author is willing to draw is that the correlation between the average failure rate and the temperature in this case is very small or non existing.

If the failures are due to a sub population[6] then, after some time, the failure rate will become even smaller at high temperature than it is at low temperature. If the average failure rate is measured over a long enough time there should not be any difference between devices run at high temperature compared to a lower temperature in a true sub population case.

RELIABILITY ENGINEERING IN THE FUTURE

It can be of interest to review the trends in quality and reliability over the past years in order to speculate about the future. It looks as we are seeing something like quantum steps in the procedures for management of quality and reliability.

Incoming inspection

In the 70's and beginning of the 80's we used to run 100% incoming inspection of components with lots of equipment and people involved.

Further improved quality provided a base for accepting Ship to Stock or Ship to Line as an incoming procedure. The work and costs for incoming inspection has decreased greatly over this period.

Qualification testing

In the 60's and 70's the quality of integrated circuits was low enough to make traditional qualification testing meaningful in some sense. 38 components could be used to verify if the quality was good or not.

In the 90's it has been virtually impossible for a single customer to run qualification tests that verifies the quality levels he is looking for. Therefore many companies today are asking for summaries of all the reliability tests performed by the vendors on a single device type or on a whole component family.

New challenges for the reliability engineer

The trend is to avoid failures rather than to catch them at the end of a line. One area that needs the knowledge of a Reliability Engineer is the Design Department. Design rules and manufacturing procedures need to be coordinated and reviwed in order to make sure that reliability is designed into the product already from the beginning.

An other area is within Life Cycle Cost management. As the need for quality and reliability information from the production and field service is growing, there will be a need for more of reliability analysis in the future.

Increased reliability implies that we need to search for reliability information where the large numbers are. That means that component qualification data must be retrieved from the manufacturer and reliability data must be retrieved from the field.

Quality information gathered from production and field service must be used in the continuous improvement process that most companies are following today.

The driving force for improvements is product economy. If the reliability engineer combines his work with the knowledge of failure and test costs it will be possible for him to give management the type of information that they need and are asking for.

CONCLUSIONS

This overview has reviewed a number of issues related to the reliability of electronics and has also presented different ways to perform field data analysis. Some of the conclusions are:

- A new model for field data analysis by the use of optimised parametric fitting has been presented

- The traditional reliability prediction method using the component part count method has been found to have very low correlation with real field data.

- The failure rate of electronics is generally not constant by time but decreasing. One exception seems to be deeply screened DRAMs that show a constant failure rate.

- The failure rate of microcircuits seem to decrease by increasing complexity as a result of continuously decreasing defect densities and effective screening procedures.

- The reliability of plastic encapsulated microcircuits seems to be even better than the so called hermetic CERDIPs that earlier were dominating.

- The failure rate as such does not appear to have a clear temperature dependence. More relevant is to talk about the temperature dependence of the life distributions involved.

- The future job for a reliability engineer is within the design department as a Reliability Designer or in Life Cycle Cost (LCC) management to analyse production and field quality data as a part of the total product improvement management.

ACKNOWLEDGEMENTS

The non parametric field statistics analysis tool that was used extensively during this study was developed by Mr Per Ericsson of our component organisation.

The method of parametric optimisation for the field study analysis of limited data was developed by Mr Patric Oscarsson within our Systems Performance organisation.

The 'ultimate hermeticity test' was developed by Mr Bengt Jarnerud at ELLEMTEL together with the author at the end of the 70's .

REFERENCES

1. J. Moltoft, 'Reliability engineering based on field information - the way ahead'. Quality and Reliability Engineering International, vol 10, No 5, pp 399 to 410, 1994 . (gives information on the non parametric field analysis method)

3. Ö. Hallberg, 'Qualification of 1MDRAMs', Ericsson Telecom AB internal report ETX/TT/GT-90:025

3. G. Peattie, 'Quality control for ICs', IEEE Spectrum, October 1981, p 96. (reliability of ICs since 1974)

4. Ö. Hallberg, 'NMOS voltage break-down characteristics compared with accelerated life tests and field use data'. IEEE/PROC.IRPS, 1981, pp 28-32 (gives field reliability information on 4-16k DRAMs)

5. Ö. Hallberg, B. Eriksson, R. Francis, R. Hjortendal, L.I. Lindberg and B. Sävström, 'Hardware reliability assurance and field experience in a telecom environment'. Quality and Reliability Engineering International, Vol. 10, pp 195-200 (1994). (gives information on some field statistics and reliability assurance procedures)

6. Ö Hallberg, 'Failure-rate as a function of time due to log-normal life distribution(s) of weak parts'. Microelectronics and Reliability, Vol. 16, pp. 155 to 158, Pergamon Press, (1977) (gives information on the calculation of failure rates using log-normal sub populations)