Microprocessor Reliability Performance as a Function of Die Location for a 0.25μ, Five Layer Metal CMOS Logic Process

Walter Carl Riordan, Russell Miller, John M. Sherman*, Jeffrey Hicks**
Intel Corporation
4500 South Dobson Road, Chandler, AZ 85248
E-mail: walter.riordan@intel.com, russell.miller@intel.com,
john.m.sherman@intel.com, jeffrey.hicks@intel.com

*4100 Sara Road, Rio Rancho, NM 87124
**2200 Mission College Blvd., Santa Clara, CA 95052

ABSTRACT

In this paper, we present the results of multiple correlations between reliability (infant mortality and other reliability metrics) and yield on a die level basis for an advanced microprocessor fabricated using a 0.25μ, five layer metal CMOS logic process. Traceability information was programmed into each unit; investigated were infant mortality of edge die verses center die, effects of unusual sort yield signatures on infant mortality, alternating row effects, and the sources of variability of burn in failures.

The model that reliability defect density is proportional to yield defect density was found to be in excellent agreement with experimental data over a wide range of yield values. The x-y die position yield was found to be an excellent predictor of infant mortality. The variation in infant mortality from wafer to wafer was found to be twice the lot to lot variation, consistent with the large number of single wafer processing tools used on advanced fabrication processes. Because the traceability information is part of the standard manufacturing flow this analysis was performed using a very large, 1 million unit sample size.

Die near the edge of the wafer were found to have worse reliability than those near the center; certain die locations were particularly poor. Unusual yield signatures at wafer sort often showed the same map of failures in burn in. The level of resolution possible from a die level analysis also allowed us to identify specific tools and interactions between tools in the fabrication process which were responsible for reliability failures.

1. INTRODUCTION

Correlation between yield and reliability has been demonstrated in prior works[1-5], but the analysis was performed with limited sample size and/or on a lot level basis. In this work the traceability feature allowed analysis to be performed on a die level basis, providing visibility into many more subtle effects such as systematic edge reliability failures. Many of these effects had been previously suspected but had not been validated with data. Moreover, because unit level traceability is part of the standard high volume manufacturing flow the analysis could be performed on a very large data set of over 1 million units, capturing a great deal of process variability which cannot normally be seen in typical reliability evaluations of limited sample size.

The traceability information was programmed into the microprocessor during wafer sort (also referred to as "wafer probe") through the use of salicided polysilicon agglomeration fuses[6]. Each die was uniquely identified with lot number, wafer number, and die location on the wafer. Die were then assembled into packaged units and burned in. The traceability information was read as part of the standard post burn in manufacturing test program and uploaded into the production database.

The material studied in this work was manufactured on a .25μ generation CMOS logic technology. The active devices on this process are fabricated in dual wells with shallow trench isolation, complementary-doped gates and source/drain regions with selective Ti-salicide to reduce resistance. There are five layers of metalization employing a Ti/AL-Cu/Ti/TiN stack, and tungsten filled plugs provide connections between layers. [7]

2. FUSE TECHNOLOGY

The traceability element in the microprocessor is a PROM arraty of fuses defined in the Ti-salicide topped polysilicon, the same layer used for the gate electrode of current Intel CMOS logic processes. The fuse can be electrically programmed through agglomeration of the salicide with current pulses, resulting in a controlled increase in resistance with no collateral damage to surrounding features. The programming can be effected at wafer sort or any subsequent electrical test step. Key advantages of this element over alternative approaches are that it requires no modification to the standard Silicon process, can be programmed at typical burn in voltages, and it is more readily scaleable. Extensive characterization has demonstrated high programming yields (fallout in the low 100s of DPM), and immeasurably low failure rates in biased and thermo-mechanical reliability testing.

3. DATA TYPES: ENGINEERING FLOW VERSES NATURALLY OCCURRING

Two sources of burn in data were used in this work: 1) engineering data and 2) naturally occurring data (also referred to as high volume manufacturing or "HVM" data). The engineering data come from the

0-7803-5220-3/99/$10.00 ©1999 IEEE

IEEE 99CH36296. 37th Annual International Reliability Physics Symposium, San Diego, California, 1999

typical material flow used to generate infant mortality data: wafer sort, packaging, unit level test, burn in, and post burn in test. The HVM data comes from the typical material flow used in high volume manufacturing: wafer sort, packaging, burn in, and post burn in test. The only difference between the engineering data and the HVM data is the absence of the unit level test prior to burn in.

There are advantages and disadvantages to each type of data. The pre-burn in unit level test in the engineering flow eliminates any possible miscorrelation between the wafer level sort test and the post burn in test, and assures that all failures detected after burn in are true infant mortality failures. The disadvantage of this flow is the cost: because this flow is not used in manufacturing, the amount of data that can be gathered is minimal, and capturing all the sources of variability is difficult.

The chief advantage of the HVM flow is the tremendous amount of data that can be gathered. Because this data is generated during manufacturing, the data is effectively "free". The data set is large enough to capture the numerous sources of variability that can affect reliability. For example, reliability data can be gathered on multiple fab runs, equipment sets, testers, work shifts, work weeks, preventative maintenance cycles, and fabrication facilities. The chief disadvantage of data from the HVM flow is that some failures to the post burn in test may have been initial failures (i.e. time=0) which would have failed to a unit level test prior to burn in. In addition, some failures may be invalid failures due to an error in testing, such as an open or shorts failure from a mis-socket.

To best filter our true reliability failures from initial failures when using HVM data, the following procedure was normally used:

1. The wafer sort test and the post burn in test use identical functional test patterns. The only differences between the two test programs are parametric: frequency, voltage, and temperature.
2. Only functional failures are included in the post burn in, unit level data set. Parametric failures for speed, standby current, a/c timings, input/output voltage levels, and input leakage are excluded.
3. Opens and shorts failures are excluded from the post burn in, unit level data set.

One of the significant results of this work is the amount of reliability information able to be gathered from naturally occurring HVM data through the use of the unit level traceability.

4. Typical Wafer Map of Burn in Failures

Figure 1 is a typical wafer map of failures seen in the post burn in test. The data in this figure comes from approximately one million packaged microprocessor units. The units were processed using the standard high volume manufacturing (HVM) flow described above, meaning no electrical test before burn in was performed other than the wafer level sort test. All failures are included in the map. None of the filtering described in section 3 above was performed.

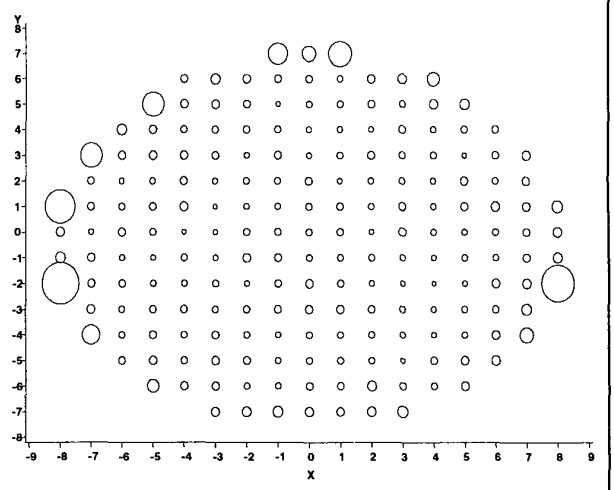

Figure 1: Wafer map of fractional fallout in burn in as a function of die location. The larger the bubble, the more burn in failures at that location. Certain edge die are particularly bad.

There are two items of note in the figure. First, more burn in failures come from die near the edge of the wafer than from those near the center. Since at wafer sort the edge of the wafer typically yields worse than the center, this has long been suspected. We are now able to validate this model with burn in data from a very large sample. Second, there are eight die locations in the figure (three at the top, four in the left or upper left, and one in the right) which have significantly more failures than the remainder of the wafer, with some locations have 8 times the failures as the wafer average. These die have become good candidates for failure analysis.

Some of the reasons for why edge die can have worse reliability than center include film deposition characteristics, plasma etching non-uniformity, wafer polishing effects, partial field lithography effects, or mechanical clips from processing equipment. For example, an incomplete metal etch at the edge of the wafer can cause metal filaments, a known burn in failure mechanism.

One of the implications of the failure map is that any process change made to improve edge die yield must consider the effect on reliability as part of the change. In addition, an application requiring extremely high reliability could consider inking out edge die altogether.

5. Reliability vs. Yield Model

Reliability has been correlated to yield in prior works [1-5]. The basic theory behind this correlation is that the same defects responsible for yield failures are also responsible for reliability failures. For a well designed process and product, intrinsic wearout failure mechanisms should not occur during the useful life of the product. Thus, the majority of failures seen in the field should be random failures due to process defect density and random "out of control" processing incidents [2].

The basic model relating reliability to yield is that reliability defect density is proportional to yield defect density. The argument goes as follows: for a total defect population, the larger defects result in yield

failures, smaller defects result in product field failures, and the smallest defects result in no failures at all [1]. This is illustrated mathematically as follows.

In this paper, a simple Poisson equation is used to model yield. For the Poisson model yield **Y** is given by

$$Y = e^{(-Dy \times A)} \qquad [1]$$

with **Dy** the yield defect density and **A** the area of the integrated circuit.

The "reliability yield" **Yr** is the fraction of devices surviving the post burn in test of the HVM flow described in section 3. In this paper, a modified Poisson equation is used to model reliability yield **Yr**

$$Yr = M \times e^{(-Dr \times A)} \qquad [2]$$

with **Dr** the reliability defect density and **M** the maximum possible yield fraction, that is, the maximum yield after burn in for material with zero reliability defects. The factor **M** is needed because of the nature of the data used in the analysis. For data obtained with the HVM flow the only electrical test prior to burn in is the wafer level sort test. Although the sort test and unit level post burn in test are similar, they are not identical; one difference is that they are performed at different temperatures. **M** is the yield to the post burn in test program as if no burn in was performed, and represents how effective the sort program is in capturing defects.

As stated earlier, reliability defect density is modeled as a fraction of the yield defect density. The proportionality constant is the reliability to yield ratio α

$$Dr = \alpha \times Dy \qquad [3]$$

The parameter α represents the nature of the manufacturing process defects and the ability of the production test program to screen these defects. A change in the nature of defects can change α; for example, a process change could create or eliminate a defect mechanism. Similarly, a change in the test program can change α; for example, the addition or elimination of a high voltage stress test.

Combining 1, 2, and 3 give the following relation between the reliability yield **Yr** and the test program yield **Y**

$$Yr = MY^{\alpha} \qquad [4]$$

$$\ln(Yr) = \alpha \ln(Y) + \ln(M) \qquad [5]$$

Thus there is a linear relationship between the logarithms of the reliability yield and post burn in test program yield, and the parameters α and **M** can be determined from the slope and intercept of a plot of ln(**Yr**) verses ln(**Y**)

$$\alpha = \text{slope} \qquad [6]$$

$$M = e^{(\text{intercept})} \qquad [7]$$

Plots of reliability after burn in verses the wafer sort test yield for a one million-unit sample of microprocessors are found in the next section. The results following indicate excellent agreement with the above model.

6. Reliability vs. Yield for a One-Million Unit Sample Size

Die level traceability allows us to examine the relationship between sort yield and reliability in much finer detail. Since traceability gives precise information on the origin of every unit in an assembly lot, it allows much better data integrity when doing a correlation. Going beyond the lot level, the data indicates wafers that yield poorly at sort relative to other wafers in a lot also have degraded reliability. Die surrounded by many failing die at sort also have lower reliability relative to die surrounded by many passing die.

Using a sample size of ~1 million units processed with the HVM flow, burn in failures (filtered as described in section 3) were compared to each of four sort yield based predictor variables: lot level yield, wafer level yield, x-y die location yield, and local region yield of the wafer. The predictors are described below.

1. Lot Level Yield is the sort yield for the lot from which the unit was taken.

2. Wafer Level Yield is the sort yield for the wafer from which the unit was taken.

3. X-Y Die Position Yield is the sort yield within the lot for units taken from the same die position on the wafer. For example, in a 10 wafer lot with 6 yielding die at x-y location (-2, 3), the x-y die position yield at this location is 6/10=60%.

4. Local Region Yield is the sort yield of the die surrounding the unit in question on the wafer. Specifically, it is the yield of the 12 nearest die weighted by distance per the following algorithm:

		T		
	D	N	D	
T	N	**X**	N	T
	D	N	D	
		T		

X is the unit in question
n = mean yield of 4 die marked N
d = mean yield of 4 die marked D
t = mean yield of 4 die marked T

$$\text{local region yield} = \left(n + d/\sqrt{2} + t/2\right) \Big/ \left(1 + 1/\sqrt{2} + 1/2\right)$$

Figures 2 through 5 are graphs of reliability yield **Yr** in burn in verses each of the above yield predictors. Each data point on the graphs includes a minimum of 30 lots (typically many more) to minimize sampling error and ensure the data is representative of the true reliability at that point. Data points not meeting this criterion were excluded. Error bars on the graph represent one sigma. Each graph is plotted on a different scale on the horizontal (yield) axis to best represent the data. The range of yield values is also different for each graph: as one would expect the range of yield values is different for each of the predictors. Note that the same 1 million unit data set is used

for each figure; the data is recalculated for each of the different sort yield predictors described above.

As can be seen in figures 2 through 5, each of the predictors fits the model of equation 4 extremely well with correlation coefficient values (R^2) greater than 0.8. The value obtained for the reliability to yield ratio α ranged from ~0.01 to ~0.02, remarkably consistent for each of the predictors. In practical terms, this means that the reliability defect density is about 1% of the yield defect density, a result consistent with that obtained by Shirley[1] on a much more limited data set.

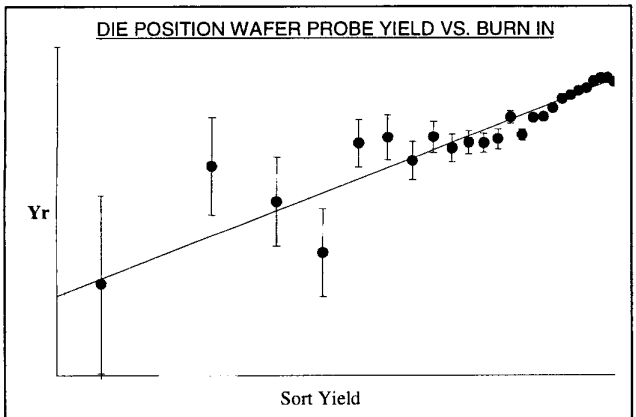

Figure 2: Reliability yield **Yr** in burn in verses X-Y die position yield. In this graph, the X-Y die position yield ranges from <5% to 100%. (α=0.0094 +/- 0.0008; sample size 993,877 units)

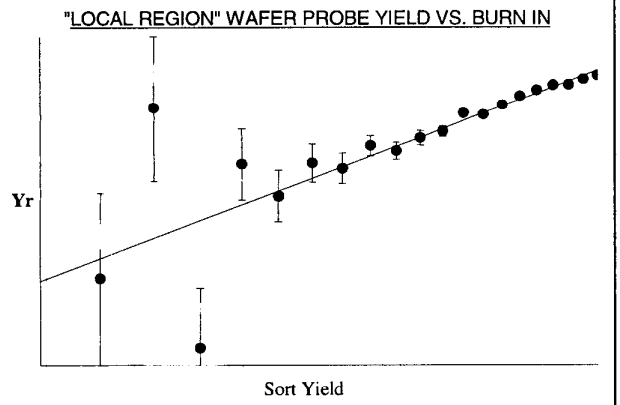

Fig 3: Reliability yield **Yr** in burn in verses local region yield. (α=0.0161 +/- 0.0016, sample size 993,276 units)

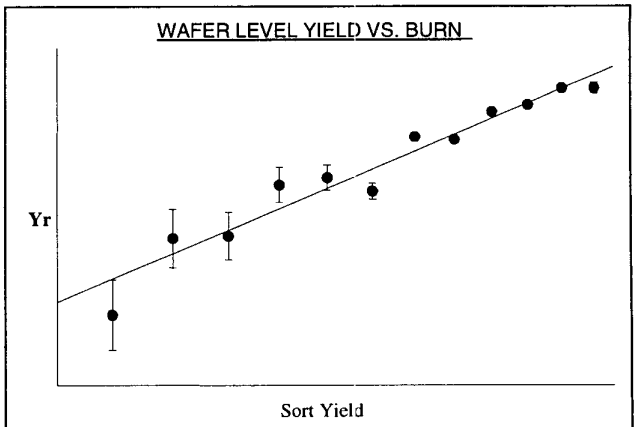

Fig 4: Reliability yield **Yr** in burn in verses wafer level yield. (α = 0.0217 +/- 0.0017, sample size 982,819 units)

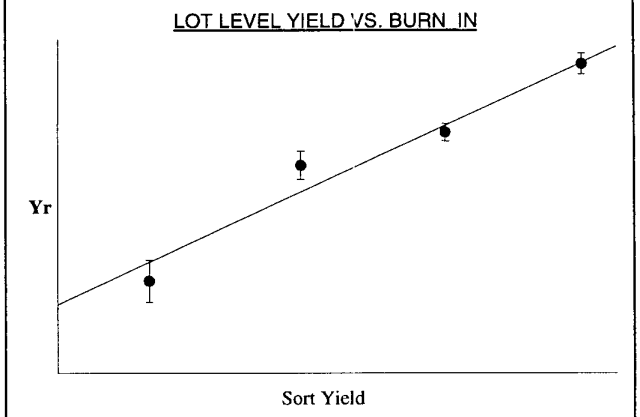

Fig 5. Reliability yield **Yr** in burn in verses lot level yield. (α = 0.0183 +/- 0.0029, sample size 928,207 units)

Two items are worthy of special note:

1. The reliability to yield model can be applied down to the die level. In other words:

   a) Low yielding die locations have degraded infant mortality. For example, certain edge die locations often yield very poorly; yielding die from these locations will have degraded burn in performance.

   b) Low yielding regions of a wafer have degraded reliability when compared to the rest of the wafer. There will be an example of this in section 8.

   c) Low yield wafers have degraded reliability when compared to higher yielding wafers

   d) Low yield lots have degraded reliability when compared to higher yielding lots.

2. The model was validated over a very wide range of yield values (<5% to 100%) and on a very large sample size. This gives very high confidence in the fundamental model.

7. SOURCES OF VARIABILITY OF BURN-IN FAILURES

Infant mortality performance in general varies on a lot level basis, with some lots having very few failures and other lots more, giving an overall process average. Using traceability information, wafer to wafer variation of burn in fallout within the lot was also evaluated. The wafer to wafer variability within a lot was found to be twice as large as the lot to lot variability. This result was determined using standard variance components techniques.

Variance components analysis is used to partition nested sources of variation. For example, a particular fab process may have lot-to-lot, wafer-to-wafer, and within-wafer variation. In this case, variance components analysis starts from the die level data and attempts to assign a unique value to each component of variation. In doing so, the procedure must separate out sampling error. For example, if a process has no inherent lot to lot variation, it will still generally produce different lot averages due to wafer to wafer variation.

In working with data of in the form of burn in data, variance components analysis gives approximate results for two reasons: the distribution is non-normal (though a normal probability transformation can nearly correct this), and the failures are discrete events, with many wafers having no failures.

Using the same 1 million-unit sample from section 6, a normal probability transform was done on the data using the SAS PROC RANK procedure. The variance components were then computed using the SAS PROC VARCOMP procedure. The result was that the *wafer to wafer variation in infant mortality was twice the run to run variation.*

As a further check of this result the actual distribution of wafer level fallout was compared to that of a simulated distribution with the same lot to lot variability. The concern was with sampling noise. Because there are a finite, small number of die available on one wafer (compared to an entire run), and because the data is attribute data (i.e. pass/fail type data) with very few failures on one wafer, the data is highly granular. One expects a large variation solely due to sampling.

The simulated data consist of failures distributed randomly throughout each lot. The result is given in figure 8. The actual distribution is seen to have many more wafers with much higher fallout than would be expected if wafer to wafer variation were truly negligible. The conclusion is that this is a true wafer to wafer variation and not due to sampling.

A possible explanation for the wafer to wafer variability exceeding the lot to lot is that for advanced processes single wafer fabrication tools are predominantly used; there is very little batch processing remaining. Effectively, one "batch" becomes a single wafer. Product qualification tests should comprehend this variability result.

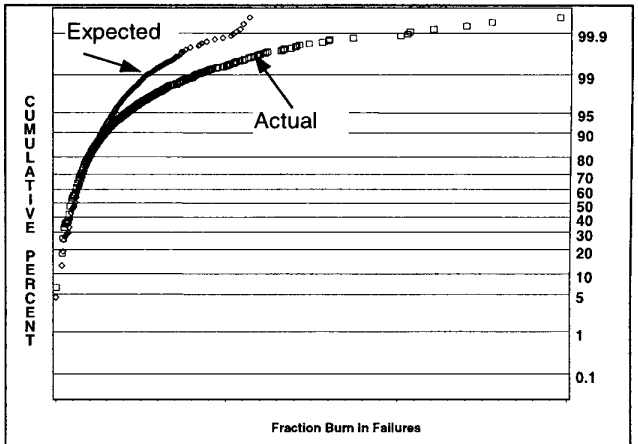

Fig 8: The curve on the left shows the expected distribution of burn in failures on a wafer if the source of variation is solely due to sampling noise. The curve on the right shows the actual distribution. The conclusion is that wafer level variability cannot be explained by sampling noise alone; there is a true wafer to wafer variation is burn in reliability.

8. CASE STUDIES

Several examples of failures detected with use of the traceability feature are now presented. The examples below generally involve interactions of multiple process variables, requiring a large data set to see the effects. The traceability feature provided us with an entire level of process "debug" enabling us to find problems which simply could not be detected in process development due to the smaller sample sizes and smaller equipment sets.

One common problem faced in manufacturing is an unusual, non-random failure patterns at sort, such as a "gross fail area" where a particular region of the wafer yields significantly different (worse) than the remainder of the wafer. Does such a failure pattern indicate a reliability problem in burn in? Our analyses found that unusual failure patterns at sort frequently, but not always, show the same failure pattern in burn in.

Metal Etch Defects

A failure pattern was noticed at wafer sort consisting of a large number of failures in lower left and upper right region of the wafer, as shown in fig. 9. The source of the defect was unknown, as was the reliability impact. The lot was assembled and processed using the engineering flow described in section 3: unit level test (t=0) was performed, and units received 12 hours burn in with readouts at 3 hours and 12 hours.

The map of failures seen in the t=0, 3, and 12 hour readouts is given in figure 10. Notice that the failure regions detected at sort (lower left and upper right) "grow" as additional failures occur. Failure analysis found the root cause to be residual Titanium left after metal etch. The defect was traced to a metal etcher (fig. 11).

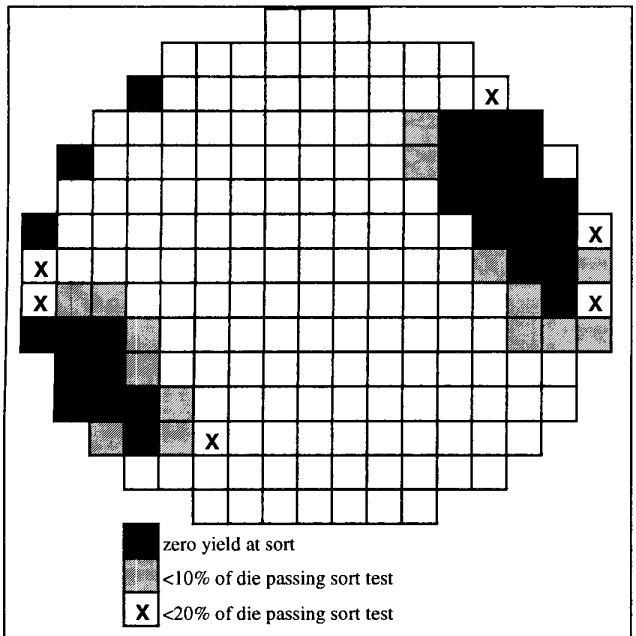

Figure 9 – Sort yield map showing two large, non-random failure regions ("Gross Failure Areas" or "GFA"s) at upper right and lower left of wafer. There were few yielding die in these two regions of the wafer. The remainder of the wafer yielded normally.

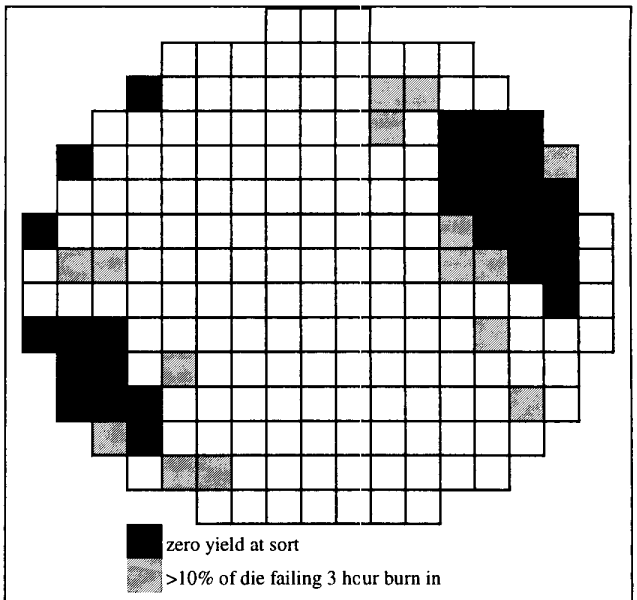

Figure 10b - 3 hour burn in results for the lot from figure 9. Die locations around the zero yield regions at sort have a high failure rate.

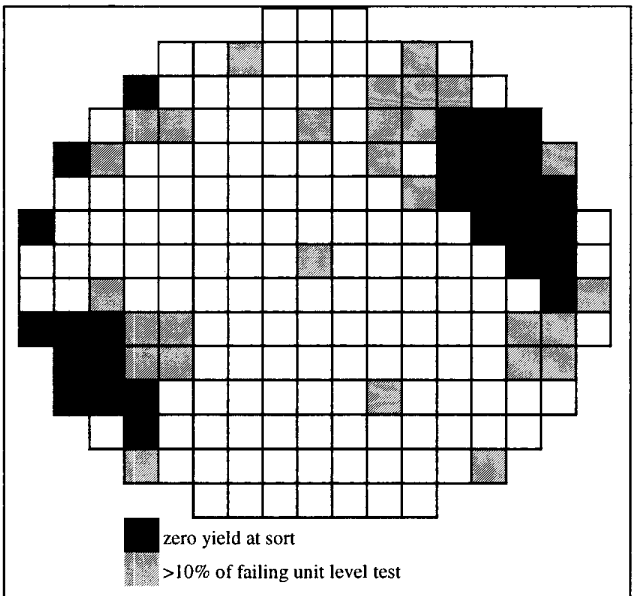

Figure 10a - Unit level test results after the lot in figure 9 was assembled into packages and tested before burn in stress. Large fractions of the die surrounding the large failure regions seen at wafer sort later failed the unit level test. The remainder of the wafer yielded normally.

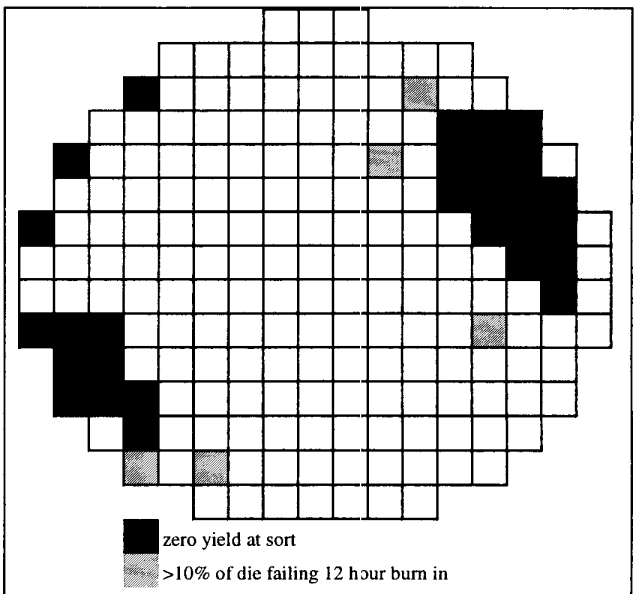

Figure 10c - 12 hour burn in results for the lot from figure 9. Die locations around the zero yield regions at sort show a high failure rate.

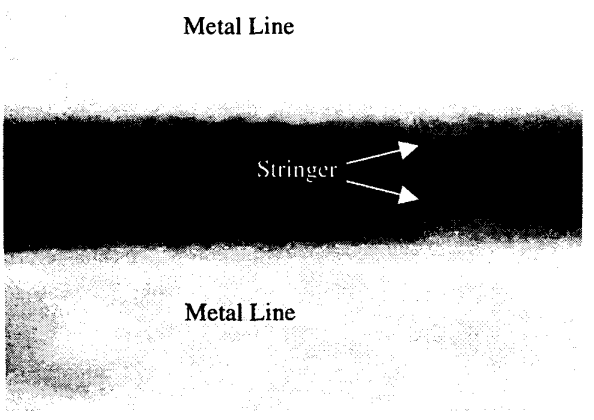

Fig 11: Top down SEM of burn in fail showing Ti filament between the two metal lines responsible for the failures seen in figures 9-10.

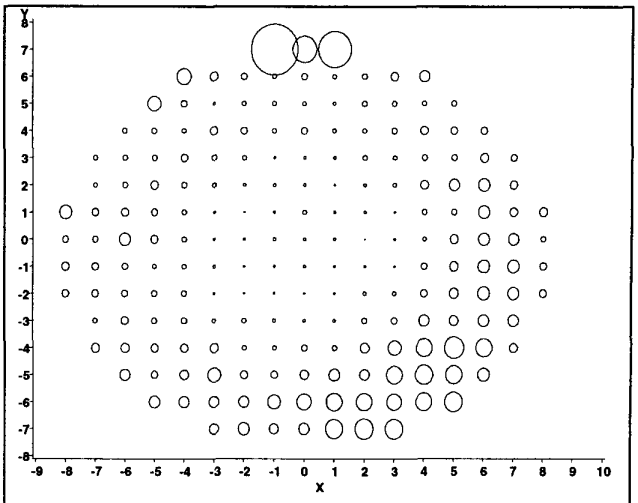

Figure 12: Wafer map of fraction of failures at sort on material affected by a process problem. Note the non-random crescent of failures in lower right portion of the wafer.

Via Defects

A process excursion was detected at wafer sort which had a "lower right crescent" failure signature. Figure 12 shows the yield wafer map from wafer sort; note the characteristic large number of failures occurring in the lower right quadrant of the wafer. Figure 13 shows the map of units failing post burn in: the same characteristic failures failure pattern in the lower right of the wafer is seen.

In this example, the burn in data in figure 13 was obtained using the HVM flow described in section 3. Similar results were obtained by processing material using the engineering flow, although since fewer units can be processed there were fewer failures and the pattern was not as obvious. Failure analysis on failures from wafer sort, burn in failures from the HVM flow, and burn in failures from the engineering flow all found the same via defect.

This example points out the tremendous utility of the traceability capability when used with naturally occurring, high volume manufacturing data. A separate engineering experiment was not necessary; the via defect could have been found using only the HVM data.

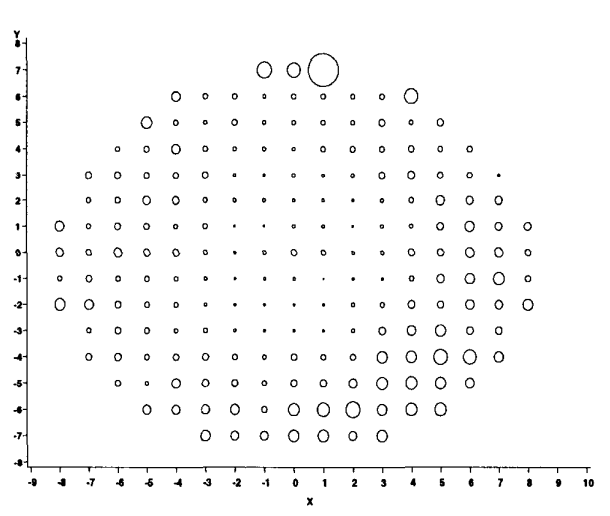

Figure 13: Wafer map of fraction of burn in failures for the material of figure 12. Note same crescent of failures, implying relationship between sort yield and burn in for this fail mechanism.

Die Cracking

Die cracking was found to occur on assembled units during the cold temperature functional test in the manufacturing flow. Failure analysis determined the root cause of the die crack was a particle on the backside of the die: the particle initiated a crack, and the cold test essentially performed a short temperature cycle stress, propagating the crack. The traceability data from these units found all failures coming from specific die locations on the edge of the wafer. From this unique signature, the source of the defect was rapidly traced to a processing tool.

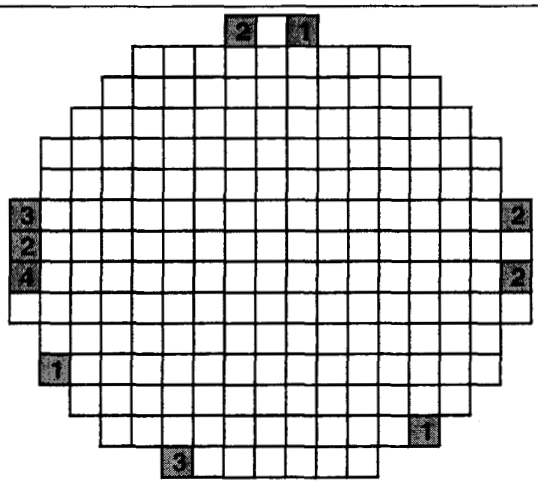

Fig 14: Wafer map showing location on wafer of units which failed for die crack. The number is the quantity of die crack failures from the die location.

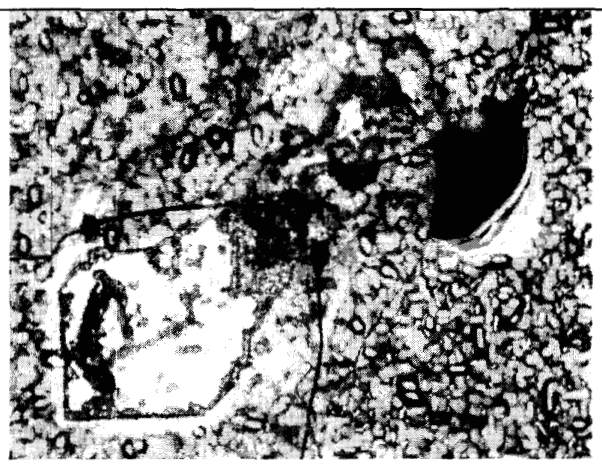

Fig 15: The aluminum particle that was the source of the die crack failure. The particle was found to originate from a dielectric film deposition tool.

Mixing in Assembly/Test

One advantage of unit level traceability is that the engineer is able to verify that die from the correct wafers and fab lots were assembled into the correct assembly lots. Occasional, unintentional mixing of die from one fab run with another fab run was identified, and the wafer assembly process was improved to eliminate this mixing.

9. Failure Mechanisms and Changes in α

As demonstrated above, defect reliability is well correlated to die yield for a well-controlled Si process, supporting the 'conventional wisdom' that improvement in die yield has direct and predictable benefits for reliability. The corollary is that any departures observed from this correlation are a concern. Increases in burn in fallout relative to die yield can be indicative of a change in the nature of latent defects or the introduction of a new failure mode or highly non-random defect. These are typically the result of subtle process marginalities, often confined to a small subset of the process tool population and/or limited periods in time. When the die yield impact is small, such defects may elude detection or diagnosis at the wafer level.

Traceability makes it possible to detect cases of elevated reliability fallout and abnormal, systematic across wafer spatial fail patterns at the component level even when they are isolated to specific Fab lots or wafers. Early detection and accelerated analysis allows such problems to be fixed before they become major excursions and ensures reliability impact is more fully comprehended in process optimization.

Below are two case studies for lots which yielded normally at sort but which had high fallout in burn in. This suggested a change in the defect characteristics and a corresponding increase in α, the reliability to yield ratio.

Metal Stringers

Burn in fallout somewhat higher than anticipated from the sort yield was encountered on an SRAM test vehicle during development of a new (.25μ) technology. Elevated numbers of single bit failures with node to node test mode signatures were responsible. Traceability showed that the failures were highly preferential to the wafer edge (Fig 16). Failure analysis identified the root cause as stringers of residual Ti shunt layers at the first metal layer, local to contacts (Fig 17). This type of defect whilst superficially very similar to a 'random' conducting particle has a highly non-random size (determined by etch and lithography parameters) and position (always at the metal edge adjacent to a contact). In certain regimes this defect resulted in much higher proportion of defects being latent than is characteristic of the baseline process--mathematically modeled as a change in α. The edge intensive fingerprint of this failure mode in burn in enabled rapid feedback confirming the effectiveness of modifications in the associated process modules.

In subsequent product reliability work, traceability found an edge-weighted distribution of burn in failures suggestive of this historical metal shorting problem, but no repeating electrical signature was apparent. Additional analysis and filtering on sub-populations of the burn in failures identified a signal: burn in failures correlated to sort yield failures having a specific single bit node to node shorting signature from a small cache of the microprocessor. These failures also correlated to a certain stepper/metal etcher combination used to process the material. In this example, the traceability feature enabled corrective actions to be implemented even before failure analysis was completed. The failure analysis subsequently confirmed metal filaments at locations of minimum metal to contact spacing as the root cause.

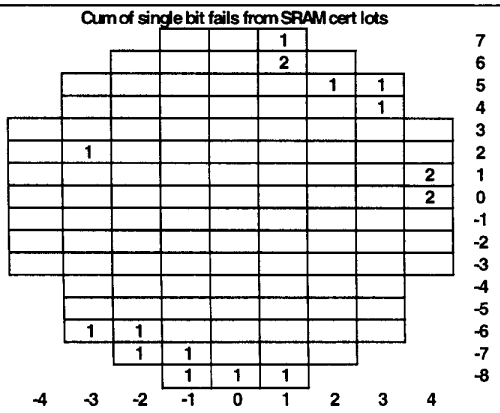

Fig 16: Single bit burn in failures from SRAM were very strongly edge-weighted; more burn in failures occurred than would be expected by the sort yield.

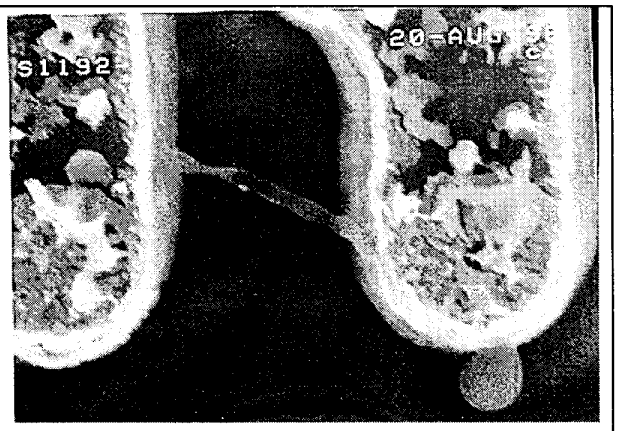

Fig 17: Single Bit SRAM failure caused by metal filament – Top View.

Ovalized Contacts

Traceability analysis on fab lots with burn in failures greater than was predicted by yield found a unique edge weighted and alternating column pattern to the failures (fig. 19). The lot histories established commonality to a specific stepper at the contact layer. Analysis of the wafer yield data found a failure pattern similar to that observed at burn in, but significantly more subtle (fig 18). Electrical analysis of the failures uncovered several repeating patterns: the cause was isolated to abnormal, ovalized contacts shorting to adjacent polysilicon. The distorted contacts were the result of marginal focus control on the affected stepper interacting with highly specific polysilicon to contact layout geometry's. This defect mode was even more systematic than the example of the metal shorts cited above. As in the prior case, there was a significant departure from the normal reliability to yield defect density ratio α.

10. PARAMETRIC EFFECTS

Alternating Column Fmax Degradation during Burn In

Due to stepper/mask characteristics, it is fairly common to see a parametric variation alternating between columns of a wafer. However, the variation in parametric *degradation* during burn in has not previously been investigated. Our analysis found that parametric degradation during burn also shows column dependence, with more severe degradation on even columns of the wafer compared to odd.

During a certification activity, a column dependence of maximum operating frequency (Fmax) degradation after burn in was discovered (figure 22). On the same wafer, units coming from even columns of the wafer degraded more than units from odd columns. The Fmax degradation is defined as the difference between a unit's Fmax value after burn in stress compared to the unit's initial Fmax value. Historically, some degradation is observed due to hot carrier effects or p-channel stability effects.

The column dependence of Fmax degradation was traced to speed path and lithography issues. Due to variation across the mask reticle field and the variation in the stepper, the limiting speed paths were different between the even and odd columns. Units from even columns had speed paths that were transistor limited and saw degradation in performance due to hot carrier effects. Units from odd columns had speed paths that were RC limited and saw little effect from transistor degradation. Hence, a difference in speed degradation was observed depending on the column a unit came from.

Standby Current Degradation during Burn In

During another certification activity, standby current was recorded for all units prior to burn in stress. The distribution of standby current before burn in was normally distributed, as is expected. After burn in a "tail" in the standby current distribution was observed. Figure 23 is a plot of the standby current after 24hrs of burn in. All three units in the tail came from the same die location on three separate wafers. It is interesting to note that this die location was a low yield location at wafer sort for pin to pin shorts.

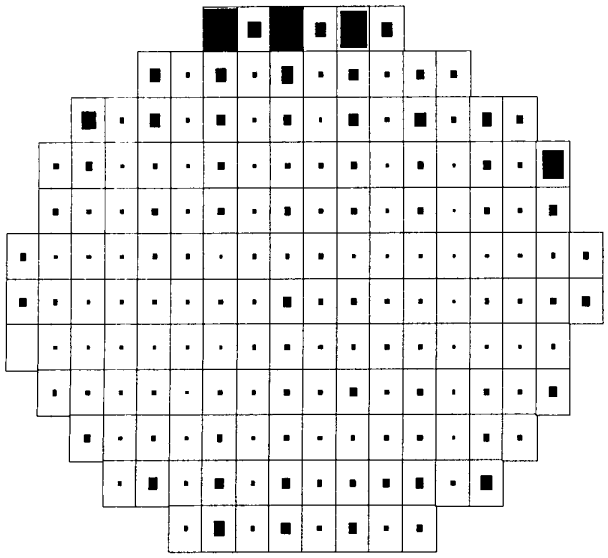

Fig 18: "Box in box" failure map for wafer sort. The larger (darker) the square in the box, the more sort failures at that location. Note that the fallout tends to be edge-weighted with some low amount of "striping" seen close to the edges.

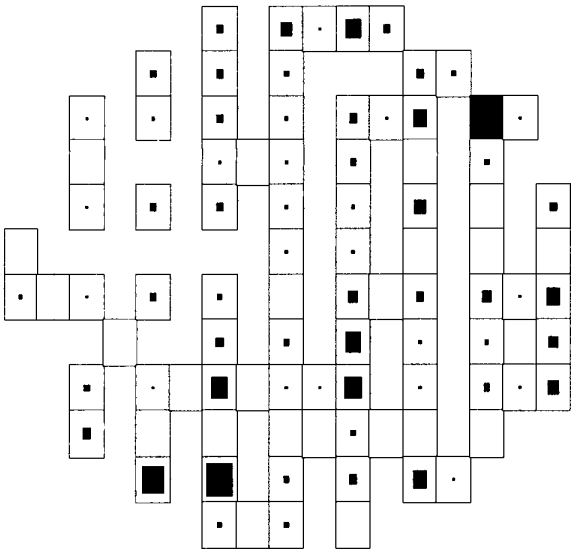

Fig 19: "Box in box" failure map of burn in failures for the material in figure 18. The larger (darker) the square in the box, the more burn in failures at that location. Strong "striping" is observed across the entire wafer, which is not seen in the sort wafer map of figure 18. A new failure mechanism was suspected because of this apparent change in α. Note that die were assembled and burned in from all sites on the wafer; the sites where there is no "box" indicates there were no burn in failures from that die location.

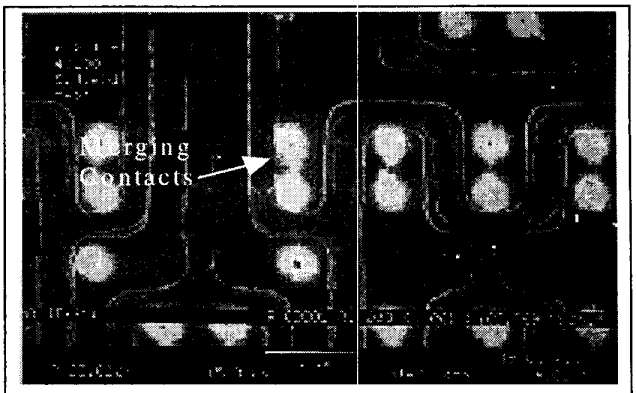

Fig 20: Failure analysis result for the burn in failures of figure 19. Etch back showing merging contact due to oval shaped contact (top view).

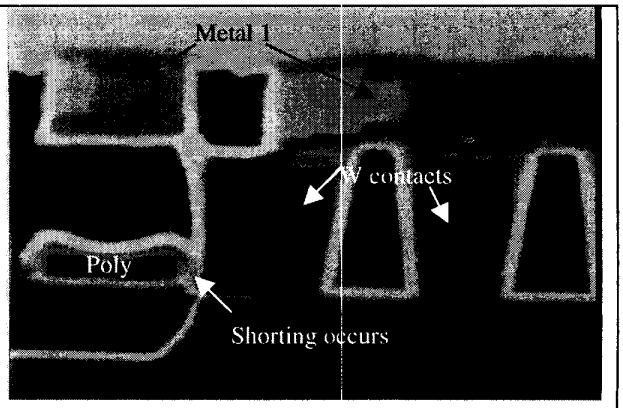

Fig 21: Failure analysis result for the burn in failures of figure 19. Cross-section shows ovalized contact shorting to adjacent poly line, and nearly shorting to adjacent metal 1.

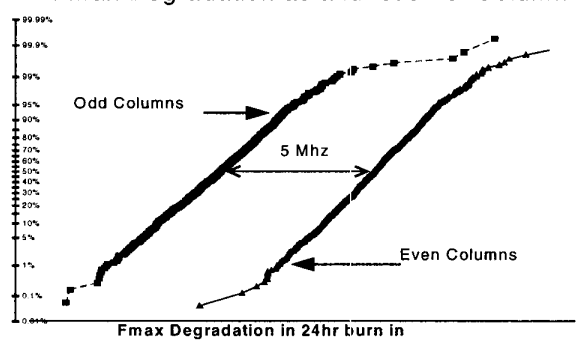

Fig 22: Probability plot of Fmax degradation (MHz) during burn in as a function of column position on the wafer. Even columns degraded an average of 5MHz greater than odd columns.

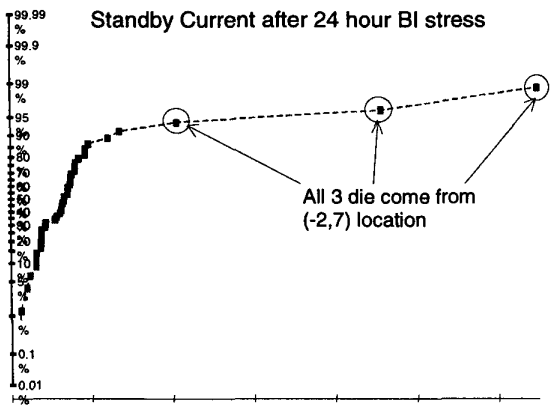

Fig 23: Probability plot of standby current after 24 hours burn in stress. The distribution is normally distributed (as expected) with the exception of the circled three units. These three units came from the same location on the wafer, at the wafer edge. The standby current distribution before burn in was normal; the three units did not have high standby current before burn in.

11. CONCLUSIONS

The relationship between burn in and wafer sort defect densities was studied with the benefit of a large population of approximately one million microprocessors incorporating an integral PROM array to provide unique die level traceability at the component level. A good, controlled correlation between burn in and wafer sort yield was observed over a wide yield range. The correlation holds for yields based on the lot, the wafer, the region of the wafer, and the die. Another way of saying this is that lower yielding regions of a wafer have worse reliability than higher yielding regions, or that lower yielding wafers have worse reliability than higher yielding wafers. As an example, a lot with a low yield "gross fail area" at wafer sort was shown to have a large fraction of die from near the low yield region fail during burn in.

Burn in failure rate was found to increase systematically at the wafer edge for the population studied, although only to the degree expected from the roll off in edge die yield. A component of variation analysis of burn in performance was performed: wafer to wafer variation in infant mortality was found to dominate the lot to lot variation by a factor of two. This is probably due to the large number of single wafer tools used in modern semiconductor manufacturing. The insight will help optimize sampling strategies in the future.

A fringe benefit of the die level traceability was simplification of component level experimental work generally obviating the need for disciplined physical segregation of experimental splits through assembly, test and stress flows.

Traceability proved to be a powerful tool for identifying isolated departures from the normal correlation between wafer yield and burn in, and for revealing characteristic within wafer patterns of such failures. This was especially true for subtle signals buried within the production burn in data set. Failure analysis established that these sub-populations were invariably differentiated by new, systematic failure modes or defect distributions. One example shown was the "ovalized" contact failure mechanism, which was detected as an alternating column pattern of failures in burn in stress.

Degradation of product performance parameters such as maximum operating frequency (Fmax) was found to have positional dependence upon the wafer. In the case of Fmax, the root cause was related to lithography equipment characteristics. The cross reticle field variation of reliability should be considered carefully in process development, and is worthy of future study.

Traceability enabled greatly accelerated diagnosis of the root causes for reliability failures. One example was die crack failures that were rapidly traced back to specific tools in the wafer fabrication process. Subtle tool-specific interactions are often highlighted, thus providing important direction for ongoing process control and optimization activities. The net result is better reliability and greater stability of reliability performance than could otherwise be achieved.

FOOTNOTES

[1] C. Glenn Shirley, "A defect model of reliability", Tutorial 3 of the International Reliability Physics Symposium, 1995.

[2] F. Kuper, J. van der Pol, E. Ooms, T. Johnson, R. Wijburg, W. Koster, D. Johnston, "Relation between Yield and Reliability of Integrated Circuits: Experimental results and Application to Continuous Early Failure Rate Reduction Programs", Proceedings of the International Reliability Physics Symposium, 1996, pp17-21.

[3] J. van der Pol, F. Kuper, E Ooms, "Relation between yield and reliability of integrated circuits and application to failure rate assessment and reduction in the one digit FIT and PPM reliability era.", Proceedings of the 7th European Symposium on Reliability of Electron Devices, Failure Physics, and Analysis, 1996, pp1603-1610.

[4] F. Evans, J. Wall, B. Hancock, P. Brusius, M. Mitchell, "Reliability in VHSIC-level 1.25μm radiation-hard CMOS IC devices ", Proceedings of the 9th biennial university/government/industry microelectronics Symposium 1991, pp. 107-112.

[5] J.G. Prendergast, "Reliability and Quality Correlation for a Particular Failure Mechanism", Proceedings of the International Reliability Physics Symposium, 1993, pp87-93.

[6] M. Alavi, M. Bohr, J. Hicks, M. Denham, A. Cassens, D. Douglas, M. Tsai, "A PROM element based on salicide agglomeration of poly fuses in a CMOS login process", IEDM, 1997.

[7] M. Bohr, S.U. Ahmed, M. Bost, T. Ghani. J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tasi, S. Yang, "A High Performance .25μ Logic Technology Optimized for 1.8V Operation", IEDM, 1996.